(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 7 September 2001 (07.09.2001)

PCT

(10) International Publication Number WO 01/64922 A2

Siena (IT). PIZZA, Mariagrazia [IT/IT]; Chiron SpA,

(51) International Patent Classification⁷: C12N 15/70, C07K 14/22, 19/00

Via Fiorentina, 1, I-53100 Siena (IT).

- (21) International Application Number: PCT/IB01/00452
- (22) International Filing Date: 28 February 2001 (28.02.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:

0004695.3 28 February 2000 (28.02.2000) GB 0027675.8 13 November 2000 (13.11.2000) GB

- (71) Applicant (for all designated States except US): CHIRON SPA [IT/IT]; Via Fiorentina 1, I-53100 Siena (IT).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): ARICO, Maria, Beatrice [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). COMANDUCCI, Maurizio [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). GALEOTTI, Cesira [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). MASIGNANI, Vega [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). GUILIANI, Marzia, Monica [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100

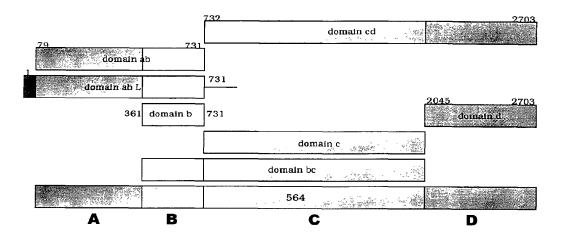
- (74) Agents: HALLYBONE, Huw, George et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS



(57) Abstract: Alternative and improved approaches to the heterologous expression of the proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. These approaches typically affect the level of expression, the ease of purification, the cellular localisation, and/or the immunological properties of the expressed protein.



-1-

HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

5

10

15

This invention is in the field of protein expression. In particular, it relates to the heterologous expression of proteins from *Neisseria* (e.g. N.gonorrhoeae or, preferably, N.meningitidis).

BACKGROUND ART

International patent applications WO99/24578, WO99/36544, WO99/57280 and WO00/22430 disclose proteins from *Neisseria meningitidis* and *Neisseria gonorrhoeae*. These proteins are typically described as being expressed in *E.coli* (*i.e.* heterologous expression) as either N-terminal GST-fusions or C-terminal His-tag fusions, although other expression systems, including expression in native *Neisseria*, are also disclosed.

It is an object of the present invention to provide alternative and improved approaches for the heterologous expression of these proteins. These approaches will typically affect the level of expression, the ease of purification, the cellular localisation of expression, and/or the immunological properties of the expressed protein.

DISCLOSURE OF THE INVENTION

Nomenclature herein

The 2166 protein sequences disclosed in WO99/24578, WO99/36544 and WO99/57280 are referred to herein by the following SEQ# numbers:

Application	Protein sequences	SEQ# herein
WO99/24578	Even SEQ IDs 2-892	SEQ#s 1-446
WO99/36544	Even SEQ IDs 2-90	SEQ#s 447-491
	Even SEQ IDs 2-3020	SEQ#s 492-2001
WO99/57280	Even SEQ IDs 3040-3114	SEQ#s 2002-2039
	SEQ IDs 3115-3241	SEQ#s 2040-2166

In addition to this SEQ# numbering, the naming conventions used in WO99/24578, WO99/36544 and WO99/57280 are also used (e.g. 'ORF4', 'ORF40', 'ORF40-1' etc. as used in WO99/24578 and WO99/36544; 'm919', 'g919' and 'a919' etc. as used in WO99/57280).

5

10

15

The 2160 proteins NMB0001 to NMB2160 from Tettelin *et al.* [Science (2000) 287:1809-1815] are referred to herein as SEQ#s 2167-4326 [see also WO00/66791].

The term 'protein of the invention' as used herein refers to a protein comprising:

- (a) one of sequences SEQ#s 1-4326; or
- (b) a sequence having sequence identity to one of SEQ#s 1-4326; or
- (c) a fragment of one of SEQ#s 1-4326.

The degree of 'sequence identity' referred to in (b) is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more). This includes mutants and allelic variants [e.g. see WO00/66741]. Identity is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters gap open penalty=12 and gap extension penalty=1. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence.

The 'fragment' referred to in (c) should comprise at least n consecutive amino acids from one of SEQ#s 1-4326 and, depending on the particular sequence, n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragment comprises an epitope from one of SEQ#s 1-4326. Preferred fragments are those disclosed in WO00/71574 and WO01/04316.

Preferred proteins of the invention are found in *N.meningitidis* serogroup B.

Preferred proteins for use according to the invention are those of serogroup B N.meningitidis strain 2996 or strain 394/98 (a New Zealand strain). Unless otherwise stated, proteins mentioned herein are from N.meningitidis strain 2996. It will be appreciated, however, that the invention is not in general limited by strain. References to a particular protein (e.g. '287', '919' etc.) may be taken to include that protein from any strain.

25 Non-fusion expression

In a first approach to heterologous expression, no fusion partner is used, and the native leader peptide (if present) is used. This will typically prevent any 'interference' from fusion partners and may alter cellular localisation and/or post-translational modification and/or folding in the heterologous host.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) no fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

The method will typically involve the step of preparing an vector for expressing a protein of the invention, such that the first expressed amino acid is the first amino acid (methionine) of said protein, and last expressed amino acid is the last amino acid of said protein (*i.e.* the codon preceding the native STOP codon).

This approach is preferably used for the expression of the following proteins using the native leader peptide: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109 and NMB2050. The suffix 'L' used herein in the name of a protein indicates expression in this manner using the native leader peptide.

15 Proteins which are preferably expressed using this approach using no fusion partner and which have no native leader peptide include: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.

Advantageously, it is used for the expression of ORF25 or ORF40, resulting in a protein which induces better anti-bactericidal antibodies than GST- or His-fusions.

20 This approach is particularly suited for expressing lipoproteins.

Leader-peptide substitution

5

10

25

30

In a second approach to heterologous expression, the native leader peptide of a protein of the invention is replaced by that of a different protein. In addition, it is preferred that no fusion partner is used. Whilst using a protein's own leader peptide in heterologous hosts can often localise the protein to its 'natural' cellular location, in some cases the leader sequence is not efficiently recognised by the heterologous host. In such cases, a leader peptide known to drive protein targeting efficiently can be used instead.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a different protein and, optionally, (b) no fusion partner is used.

5

15

20

30

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide and to introduce nucleotides that encode a different protein's leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The expressed protein will consist of the replacement leader peptide at the N-terminus, followed by the protein of the invention minus its leader peptide.

The leader peptide is preferably from another protein of the invention (e.g. one of SEQ#s 1-4326), but may also be from an *E.coli* protein (e.g. the OmpA leader peptide) or an *Erwinia carotovora* protein (e.g. the PelB leader peptide), for instance.

A particularly useful replacement leader peptide is that of ORF4. This leader is able to direct lipidation in E.coli, improving cellular localisation, and is particularly useful for the expression of proteins 287, 919 and Δ G287. The leader peptide and N-terminal domains of 961 are also particularly useful.

Another useful replacement leader peptide is that of *E.coli* OmpA. This leader is able to direct membrane localisation of *E.coli*. It is particularly advantageous for the expression of ORF1, resulting in a protein which induces better anti-bactericidal antibodies than both fusions and protein expressed from its own leader peptide.

Another useful replacement leader peptide is MKKYLFSAA. This can direct secretion into culture medium, and is extremely short and active. The use of this leader peptide is not restricted to the expression of Neisserial proteins – it may be used to direct the expression of any protein (particularly bacterial proteins).

Leader-peptide deletion

In a third approach to heterologous expression, the native leader peptide of a protein of the invention is deleted. In addition, it is preferred that no fusion partner is used.

25 Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may PCT/IB01/00452

already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

-5-

This method can increase the levels of expression. For protein 919, for example, expression levels in *E.coli* are much higher when the leader peptide is deleted. Increased expression may be due to altered localisation in the absence of the leader peptide.

The method is preferably used for the expression of 919, ORF46, 961, 050-1, 760 and 287.

Domain-based expression

WO 01/64922

5

15

20

25

In a fourth approach to heterologous expression, the protein is expressed as domains. This may be used in association with fusion systems (e.g. GST or His-tag fusions).

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove at least one domain from within the protein. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. Where no fusion partners are used, the first amino acid of the expressed protein will be that of a domain of the protein.

A protein is typically divided into notional domains by aligning it with known sequences in databases and then determining regions of the protein which show different alignment patterns from each other.

The method is preferably used for the expression of protein 287. This protein can be notionally split into three domains, referred to as A B & C (see Figure 5). Domain B aligns strongly with IgA proteases, domain C aligns strongly with transferrin-binding proteins, and domain A shows no strong alignment with database sequences. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

Once a protein has been divided into domains, these can be (a) expressed singly (b) deleted from with the protein e.g. protein ABCD \rightarrow ABD, ACD, BCD etc. or (c) rearranged e.g. protein ABC \rightarrow ACB, CAB etc. These three strategies can be combined with fusion partners is desired.

ORF46 has also been notionally split into two domains – a first domain (amino acids 1-433) which is well-conserved between species and serogroups, and a second domain (amino acids 433-608) which is not well-conserved. The second domain is preferably deleted. An alignment of polymorphic forms of ORF46 is disclosed in WO00/66741.

5 Protein 564 has also been split into domains (Figure 8), as have protein 961 (Figure 12) and protein 502 (amino acids 28-167 of the MC58 protein).

Hybrid proteins

10

20

In a fifth approach to heterologous expression, two or more (e.g. 3, 4, 5, 6 or more) proteins of the invention are expressed as a single hybrid protein. It is preferred that no non-Neisserial fusion partner (e.g. GST or poly-His) is used.

This offers two advantages. Firstly, a protein that may be unstable or poorly expressed on its own can be assisted by adding a suitable hybrid partner that overcomes the problem. Secondly, commercial manufacture is simplified – only one expression and purification need be employed in order to produce two separately-useful proteins.

15 Thus the invention provides a method for the simultaneous heterologous expression of two or more proteins of the invention, in which said two or more proteins of the invention are fused (i.e. they are translated as a single polypeptide chain).

The method will typically involve the steps of: obtaining a first nucleic acid encoding a first protein of the invention; obtaining a second nucleic acid encoding a second protein of the invention; ligating the first and second nucleic acids. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Preferably, the constituent proteins in a hybrid protein according to the invention will be from the same strain.

The fused proteins in the hybrid may be joined directly, or may be joined via a linker peptide e.g. via a poly-glycine linker (i.e. G_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more) or via a short peptide sequence which facilitates cloning. It is evidently preferred not to join a ΔG protein to the C-terminus of a poly-glycine linker.

The fused proteins may lack native leader peptides or may include the leader peptide sequence of the N-terminal fusion partner.

The method is well suited to the expression of proteins orf1, orf4, orf25, orf40, Orf46/46.1, orf83, 233, 287, 292L, 564, 687, 741, 907, 919, 953, 961 and 983.

The 42 hybrids indicated by 'X' in the following table of form NH₂-A—B-COOH are preferred:

$\downarrow A$ $B \rightarrow$	ORF46.1	287	741	919	953	961	983
ORF46.1		X	Х	Х	Х	Х	X
287	X		Х	Х	Х	Х	X
741	X	X		Х	Х	Х	Х
919	Х	Х	Х		Х	Х	Х
953	Х	X	Х	Х		Х	Х
961	X	Х	Х	Х	Х		Х
983	X	X	Х	Х	Х	X	

Preferred proteins to be expressed as hybrids are thus ORF46.1, 287, 741, 919, 953, 961 and 983. These may be used in their essentially full-length form, or poly-glycine deletions (ΔG) forms may be used (e.g. ΔG-287, ΔGTbp2, ΔG741, ΔG983 etc.), or truncated forms may be used (e.g. Δ1-287, Δ2-287 etc.), or domain-deleted versions may be used (e.g. 287B, 287C, 287BC, ORF46₁₋₄₃₃, ORF46₄₃₃₋₆₀₈, ORF46, 961c etc.).

Particularly preferred are: (a) a hybrid protein comprising 919 and 287; (b) a hybrid protein comprising 953 and 287; (c) a hybrid protein comprising 287 and ORF46.1; (d) a hybrid protein comprising ORF1 and ORF46.1; (e) a hybrid protein comprising 919 and ORF46.1; (f) a hybrid protein comprising ORF46.1 and 919; (g) a hybrid protein comprising ORF46.1, 287 and 919; (h) a hybrid protein comprising 919 and 519; and (i) a hybrid protein comprising ORF97 and 225. Further embodiments are shown in Figure 14.

Where 287 is used, it is preferably at the C-terminal end of a hybrid; if it is to be used at the N-terminus, if is preferred to use a ΔG form of 287 is used (e.g. as the N-terminus of a hybrid with ORF46.1, 919, 953 or 961).

Where 287 is used, this is preferably from strain 2996 or from strain 394/98.

Where 961 is used, this is preferably at the N-terminus. Domain forms of 961 may be used.

Alignments of polymorphic forms of ORF46, 287, 919 and 953 are disclosed in WO00/66741. Any of these polymorphs can be used according to the present invention.

Temperature

WO 01/64922

In a sixth approach to heterologous expression, proteins of the invention are expressed at a low temperature.

Expressed Neisserial proteins (e.g. 919) may be toxic to *E.coli*, which can be avoided by expressing the toxic protein at a temperature at which its toxic activity is not manifested.

Thus the present invention provides a method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.

A preferred temperature is around 30°C. This is particularly suited to the expression of 919.

10 Mutations

5

As discussed above, expressed Neisserial proteins may be toxic to *E.coli*. This toxicity can be avoided by mutating the protein to reduce or eliminate the toxic activity. In particular, mutations to reduce or eliminate toxic enzymatic activity can be used, preferably using site-directed mutagenesis.

In a seventh approach to heterologous expression, therefore, an expressed protein is mutated to reduce or eliminate toxic activity.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

The method is preferably used for the expression of protein 907, 919 or 922. A preferred mutation in 907 is at Glu-117 (e.g. Glu→Gly); preferred mutations in 919 are at Glu-255 (e.g. Glu→Gly) and/or Glu-323 (e.g. Glu→Gly); preferred mutations in 922 are at Glu-164 (e.g. Glu→Gly), Ser-213 (e.g. Ser→Gly) and/or Asn-348 (e.g. Asn→Gly).

Alternative vectors

25

In a eighth approach to heterologous expression, an alternative vector used to express the protein. This may be to improve expression yields, for instance, or to utilise plasmids that are already approved for GMP use.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which an alternative vector is used. The alternative vector is preferably pSM214, with no fusion partners. Leader peptides may or may not be included.

This approach is particularly useful for protein 953. Expression and localisation of 953 with its native leader peptide expressed from pSM214 is much better than from the pET vector.

pSM214 may also be used with: Δ G287, Δ 2-287, Δ 3-287, Δ 4-287, Orf46.1, 961L, 961, 961(MC58), 961c, 961c-L, 919, 953 and Δ G287-Orf46.1.

Another suitable vector is pET-24b (Novagen; uses kanamycin resistance), again using no fusion partners. pET-24b is preferred for use with: ΔG287K, Δ2-287K, Δ3-287K, Δ4-287K, Orf46.1-K, Orf46A-K, 961-K (MC58), 961a-K, 961b-K, 961c-K, 961c-L-K, 961d-K, ΔG287-919-K, ΔG287-Orf46.1-K and ΔG287-961-K.

Multimeric form

In a ninth approach to heterologous expression, a protein is expressed or purified such that it adopts a particular multimeric form.

This approach is particularly suited to protein 953. Purification of one particular multimeric form of 953 (the monomeric form) gives a protein with greater bactericidal activity than other forms (the dimeric form).

Proteins 287 and 919 may be purified in dimeric forms.

Protein 961 may be purified in a 180kDa oligomeric form (e.g. a tetramer).

Lipidation

In a tenth approach to heterologous expression, a protein is expressed as a lipidated protein.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.

This is particularly useful for the expression of 919, 287, ORF4, 406, 576-1, and ORF25. Polymorphic forms of 919, 287 and ORF4 are disclosed in WO00/66741.

The method will typically involve the use of an appropriate leader peptide without using an N-terminal fusion partner.

25 C-terminal deletions

In an eleventh approach to heterologous expression, the C-terminus of a protein of the invention is mutated. In addition, it is preferred that no fusion partner is used.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to mutate nucleotides that encode the protein's C-terminus portion. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

The mutation may be a substitution, insertion or, preferably, a deletion.

This method can increase the levels of expression, particularly for proteins 730, ORF29 and ORF46. For protein 730, a C-terminus region of around 65 to around 214 amino acids may be deleted; for ORF46, the C-terminus region of around 175 amino acids may be deleted; for ORF29, the C-terminus may be deleted to leave around 230-370 N-terminal amino acids.

Leader peptide mutation

5

In a twelfth approach to heterologous expression, the leader peptide of the protein is mutated. This is particularly useful for the expression of protein 919.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides within the leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Poly-glycine deletion

25

In a thirteenth approach to heterologous expression, poly-glycine stretches in wild-type sequences are mutated. This enhances protein expression.

The poly-glycine stretch has the sequence $(Gly)_n$, where $n\geq 4$ (e.g. 5, 6, 7, 8, 9 or more). This stretch is mutated to disrupt or remove the $(Gly)_n$. This may be by deletion (e.g. CGGGGS \rightarrow CGGGGS, CGS, CGS or CS), by substitution (e.g. CGGGGS \rightarrow CGXGGS, CGXGGS, CGXGGS, etc.).

-11-

This approach is not restricted to Neisserial proteins – it may be used for any protein (particularly bacterial proteins) to enhance heterologous expression. For Neisserial proteins, however, it is particularly suitable for expressing 287, 741, 983 and Tbp2. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) a poly-glycine stretch within the protein is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides that encode a polyglycine stretch within the protein sequence. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Conversely, the opposite approach (*i.e.* introduction of poly-glycine stretches) can be used to suppress or diminish expression of a given heterologous protein.

Heterologous host

10

15

25

Whilst expression of the proteins of the invention may take place in the native host (i.e. the organism in which the protein is expressed in nature), the present invention utilises a heterologous host. The heterologous host may be prokaryotic or eukaryotic. It is preferably E.coli, but other suitable hosts include Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonenna typhimurium, Neisseria meningitidis, Neisseria gonorrhoeae, Neisseria lactamica, Neisseria cinerea, Mycobateria (e.g. M.tuberculosis), yeast etc.

20 Vectors etc.

As well as the methods described above, the invention provides (a) nucleic acid and vectors useful in these methods (b) host cells containing said vectors (c) proteins expressed or expressable by the methods (d) compositions comprising these proteins, which may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions (e) these compositions for use as medicaments (e.g. as vaccines) or as diagnostic reagents (f) the use of these compositions in the manufacture of (1) a medicament for treating or preventing infection due to Neisserial bacteria (2) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria, and/or (3) a reagent which can raise antibodies against Neisserial bacteria and (g) a method of treating a

patient, comprising administering to the patient a therapeutically effective amount of these compositions.

Sequences

10

15

The invention also provides a protein or a nucleic acid having any of the sequences set out in the following examples. It also provides proteins and nucleic acid having sequence identity to these. As described above, the degree of 'sequence identity' is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the nucleic acid disclosed in the examples, preferably under "high stringency" conditions (eg. 65°C in a 0.1xSSC, 0.5% SDS solution).

The invention also provides nucleic acid encoding proteins according to the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (eg. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) etc.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1 and 2 show constructs used to express proteins using heterologous leader peptides.

Figure 3 shows expression data for ORF1, and Figure 4 shows similar data for protein 961.

Figure 5 shows domains of protein 287, and Figures 6 & 7 show deletions within domain A.

Figure 8 shows domains of protein 564.

Figure 9 shows the *PhoC* reporter gene driven by the 919 leader peptide, and Figure 10 shows the results obtained using mutants of the leader peptide.

Figure 11 shows insertion mutants of protein 730 (A: 730-C1; B: 730-C2).

Figure 12 shows domains of protein 961.

-13-

Figure 13 shows SDS-PAGE of ΔG proteins. Dots show the main recombinant product.

Figure 14 shows 26 hybrid proteins according to the invention.

MODES FOR CARRYING OUT THE INVENTION

Example 1 – 919 and its leader peptide

5 Protein 919 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```
1 MKKYLFRAAL YGIAAAILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
51 GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
101 CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
151 RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
201 HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
251 EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
301 KLGQTSMQGI KAYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
351 TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
401 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

15 The leader peptide is underlined.

The sequences of 919 from other strains can be found in Figures 7 and 18 of WO00/66741.

Example 2 of WO99/57280 discloses the expression of protein 919 as a His-fusion in *E.coli*. The protein is a good surface-exposed immunogen.

Three alternative expression strategies were used for 919:

1) 919 without its leader peptide (and without the mature N-terminal cysteine) and without any fusion partner ('919^{untagged}'):

```
25 QSKSIQTFP QPDTSVINGP DRPVGIPDPA GTTVGGGGAV YTVVPHLSLP
100 YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR RTAQARFPIY GIPDDFISVP
150 LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT HTADLSRFPI TARTTAIKGR
200 FEGSRFLPYH TRNQINGGAL DGKAPILGYA EDPVELFFMH IQGSGRLKTP
250 SGKYIRIGYA DKNEHPYVSI GRYMADKGYL KLGQTSMQGI KAYMRQNPQR
300 LAEVLGQNPS YIFFRELAGS SNDGPVGALG TPLMGEYAGA VDRHYITLGA
350 PLFVATAHPV TRKALNRLIM AQDTGSAIKG AVRVDYFWGY GDEAGELAGK
360 QKTTGYVWQL LPNGMKPBYR P*
```

The leader peptide and cysteine were omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

- 2) 919 with its own leader peptide but without any fusion partner ('919L'); and
- 35 3) 919 with the leader peptide (MKTFFKTLSAAALALILAA) from ORF4 ('919LOrf4').

```
40 MKTFFKTLS AAALALILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
50 GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
100 CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
150 RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
150 HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
250 EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
300 KLGQTSMQGI KSYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
```

-14-

- 350 TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
- 400 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*

To make this construct, the entire sequence encoding the ORF4 leader peptide was included in the 5'-primer as a tail (primer 919Lorf4 For). A *NheI* restriction site was generated by a double nucleotide change in the sequence coding for the ORF4 leader (no amino acid changes), to allow different genes to be fused to the ORF4 leader peptide sequence. A stop codon was included in all the 3'-end primer sequences.

All three forms of the protein were expressed and could be purified.

The '919L' and '919LOrf4' expression products were both lipidated, as shown by the incorporation of [³H]-palmitate label. 919^{untagged} did not incorporate the ³H label and was located intracellularly.

919LOrf4 could be purified more easily than 919L. It was purified and used to immunise mice. The resulting sera gave excellent results in FACS and ELISA tests, and also in the bactericidal assay. The lipoprotein was shown to be localised in the outer membrane.

919^{untagged} gave excellent ELISA titres and high serum bactericidal activity. FACS confirmed its cell surface location.

Example 2 – 919 and expression temperature

5

15

25

30

Growth of *E.coli* expressing the 919LOrf4 protein at 37°C resulted in lysis of the bacteria. In order to overcome this problem, the recombinant bacteria were grown at 30°C. Lysis was prevented without preventing expression.

Example 3 – mutation of 907, 919 and 922

It was hypothesised that proteins 907, 919 and 922 are murein hydrolases, and more particularly lytic transglycosylases. Murein hydrolases are located on the outer membrane and participate in the degradation of peptidoglycan.

The purified proteins 919^{untagged}, 919Lorf4, 919-His (*i.e.* with a C-terminus His-tag) and 922-His were thus tested for murein hydrolase activity [Ursinus & Holtje (1994) *J.Bact.* 176:338-343]. Two different assays were used, one determining the degradation of insoluble murein sacculus into soluble muropeptides and the other measuring breakdown of poly(MurNAc-GlcNAc)_{n>30} glycan strands.

-15-

The first assay uses murein sacculi radiolabelled with meso-2,6-diamino-3,4,5-[³H]pimelic acid as substrate. Enzyme (3–10 µg total) was incubated for 45 minutes at 37°C in a total volume of 100µl comprising 10mM Tris-maleate (pH 5.5), 10mM MgCl₂, 0.2% v/v Triton X-100 and [³H]A₂pm labelled murein sacculi (about 10000cpm). The assay mixture was placed on ice for 15 minutes with 100 µl of 1% w/v N-acetyl-N,N,N-trimethylammonium for 15 minutes and precipitated material pelleted by centrifugation at 10000g for 15 minutes. The radioactivity in the supernatant was measured by liquid scintillation counting. *E.coli* soluble lytic transglycosylase Slt70 was used as a positive control for the assay; the negative control comprised the above assay solution without enzyme.

All proteins except 919-His gave positive results in the first assay.

5

15

25

30

The second assay monitors the hydrolysis of poly(MurNAc-GlcNAc)glycan strands. Purified strands, poly(MurNAc-GlcNAc)_{n>30} labelled with N-acetyl-D-1-[³H]glucosamine were incubated with 3µg of 919L in 10 mM Tris-maleate (pH 5.5), 10 mM MgCl₂ and 0.2% v/v Triton X-100 for 30 min at 37°C. The reaction was stopped by boiling for 5 minutes and the pH of the sample adjusted to about 3.5 by addition of 10µl of 20% v/v phosphoric acid. Substrate and product were separated by reversed phase HPLC on a Nucleosil 300 C₁₈ column as described by Harz *et. al.* [*Anal. Biochem.* (1990) 190:120-128]. The *E.coli* lytic transglycosylase Mlt A was used as a positive control in the assay. The negative control was performed in the absence of enzyme.

By this assay, the ability of 919LOrf4 to hydrolyse isolated glycan strands was demonstrated when anhydrodisaccharide subunits were separated from the oligosaccharide by HPLC.

Protein 919Lorf4 was chosen for kinetic analyses. The activity of 919Lorf4 was enhanced 3.7-fold by the addition of 0.2% v/v Triton X-100 in the assay buffer. The presence of Triton X-100 had no effect on the activity of 919^{untagged}. The effect of pH on enzyme activity was determined in Tris-Maleate buffer over a range of 5.0 to 8.0. The optimal pH for the reaction was determined to be 5.5. Over the temperature range 18°C to 42°C, maximum activity was observed at 37°C. The effect of various ions on murein hydrolase activity was determined by performing the reaction in the presence of a variety of ions at a final concentration of 10mM. Maximum activity was found with Mg²⁺, which stimulated activity 2.1-fold. Mn²⁺ and Ca²⁺ also stimulated enzyme activity to a similar extent while the addition Ni²⁺ and EDTA had no significant effect. In contrast, both Fe²⁺ and Zn²⁺ significantly inhibited enzyme activity.

WO 01/64922

5

15

20

PCT/IB01/00452

The structures of the reaction products resulting from the digestion of unlabelled *E.coli* murein sacculus were analysed by reversed-phase HPLC as described by Glauner [*Anal. Biochem.* (1988) 172:451-464]. Murein sacculi digested with the muramidase Cellosyl were used to calibrate and standardise the Hypersil ODS column. The major reaction products were 1,6 anhydrodisaccharide tetra and tri peptides, demonstrating the formation of 1,6 anhydromuraminic acid intramolecular bond.

These results demonstrate experimentally that 919 is a murein hydrolase and in particular a member of the lytic transglycosylase family of enzymes. Furthermore the ability of 922-His to hydrolyse murein sacculi suggests this protein is also a lytic transglycosylase.

10 This activity may help to explain the toxic effects of 919 when expressed in *E.coli*.

In order to eliminate the enzymatic activity, rational mutagenesis was used. 907, 919 and 922 show fairly low homology to three membrane-bound lipidated murein lytic transglycosylases from *E.coli*:

919 (441aa) is 27.3% identical over 440aa overlap to *E.coli* MLTA (P46885);

922 (369aa) is 38.7% identical over 310aa overlap to *E.coli* MLTB (P41052); and

907-2 (207aa) is 26.8% identical over 149aa overlap to E.coli MLTC (P52066).

907-2 also shares homology with *E.coli* MLTD (P23931) and Slt70 (P03810), a soluble lytic transglycosylase that is located in the periplasmic space. No significant sequence homology can be detected among 919, 922 and 907-2, and the same is true among the corresponding MLTA, MLTB and MLTC proteins.

Crystal structures are available for Slt70 [1QTEA; 1QTEB; Thunnissen *et al.* (1995) *Biochemistry* 34:12729-12737] and for Slt35 [1LTM; 1QUS; 1QUT; van Asselt *et al.* (1999) *Structure Fold Des* 7:1167-80] which is a soluble form of the 40kDa MLTB.

The catalytic residue (a glutamic acid) has been identified for both Slt70 and MLTB.

In the case of Slt70, mutagenesis studies have demonstrated that even a conservative substitution of the catalytic Glu505 with a glutamine (Gln) causes the complete loss of enzymatic activity. Although Slt35 has no obvious sequence similarity to Slt70, their catalytic domains shows a surprising similarity. The corresponding catalytic residue in MLTB is Glu162.

Another residue which is believed to play an important role in the correct folding of the enzymatic cleft is a well-conserved glycine (Gly) downstream of the glutamic acid. Recently, Terrak *et al.* [Mol.Microbiol. (1999) 34:350-64] have suggested the presence of another important residue which is an aromatic amino acid located around 70-75 residues downstream of the catalytic glutamic acid.

Sequence alignment of Slt70 with 907-2 and of MLTB with 922 were performed in order to identify the corresponding catalytic residues in the MenB antigens.

The two alignments in the region of the catalytic domain are reported below:

907-2/Slt70:

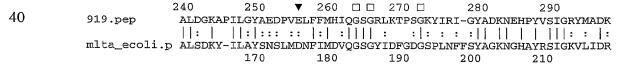
5

							
10		90	100	110	▼ 120	130	140
	907-2.pep	ERRRLLVNI	QYESSRAG	LDTQIVLGLI	EV E SAFRQYA:	SGV G AR G LMQ	VMPFWKNYIG
			:: :	: :::::	: : :	-1 1 <u>1</u> 1	: ::
	slty_ecoli		LFKRYTSGKE	IPQSYAMAIA	RQ E SAWNPKVI	KSPV G AS G LMQ	IMPGTATHTV
15		480	490	500	▲ 510	520	530
15					GLU505		
	<u>922/MLTB</u>						
		150	160 ▼	170	180	190	200
• •	922.pep	VAQKYGVPA	ELIVAVIGI e	TNY G KNT G SF	RVADALATLGI	DYPRRAGFFQ	KELVELLKLA
20		:]]]]]]:]]::]]:]]:]:]:]:]]]]]:	: : :	: :
	mltb_ecoli	AWQVYGVPP				NYPRRAEYFS	
		150	160 ▲		180	190	200
			G	LU162			
25		210	220	230	240	250	260
23	922.pep					250 HRDIWGNVGDV	260
	322.pep	:: :	· .	7	:: ::	:: :	: :
	mltb_ecoli		·! •			:: : HINLWDPV-DA	* 1 1 1 1 1 1 1
		210	220	230	240	250	260
30		220	220	250	240	250	200

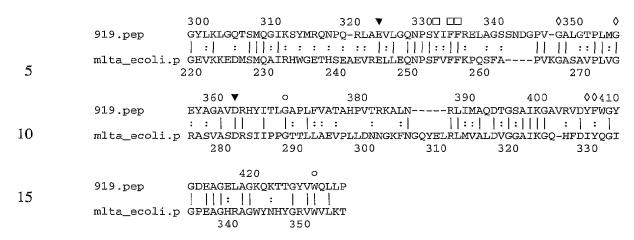
From these alignments, it results that the corresponding catalytic glutamate in 907-2 is Glu117, whereas in 922 is Glu164. Both antigens also share downstream glycines that could have a structural role in the folding of the enzymatic cleft (in bold), and 922 has a conserved aromatic residue around 70aa downstream (in bold).

In the case of protein 919, no 3D structure is available for its *E.coli* homologue MLTA, and nothing is known about a possible catalytic residue. Nevertheless, three amino acids in 919 are predicted as catalytic residues by alignment with MLTA:

919/MLTA



-18-



The three possible catalytic residues are shown by the symbol ∇ :

25

- 20 1) Glu255 (Asp in MLTA), followed by three conserved glycines (Gly263, Gly265 and Gly272) and three conserved aromatic residues located approximately 75-77 residues downstream. These downstream residues are shown by □.
 - 2) Glu323 (conserved in MLTA), followed by 2 conserved glycines (Gly347 and Gly355) and two conserved aromatic residues located 84-85 residues downstream (Tyr406 or Phe407). These downstream residues are shown by ◊.
 - 3) Asp362 (instead of the expected Glu), followed by one glycine (Gly 369) and a conserved aromatic residue (Trp428). These downstream residues are shown by o.

Alignments of polymorphic forms of 919 are disclosed in WO00/66741.

Based on the prediction of catalytic residues, three mutants of the 919 and one mutant of 907, containing each a single amino acid substitution, have been generated. The glutamic acids in position 255 and 323 and the aspartic acids in position 362 of the 919 protein and the glutamic acid in position 117 of the 907 protein, were replaced with glycine residues using PCR-based SDM. To do this, internal primers containing a codon change from Glu or Asp to Gly were designed:

Primers	Sequences	Codon change
919-E255 for	CGAAGACCCCGTCGgtCTTTTTTTATG	$GAA \rightarrow Ggt$
919-E255 rev	GTGCATAAAAAAAAGacCGACGGGGTCT	
919-E323 for	AACGCCTCGCC <u>Ggt</u> GTTTTGGGTCA	$GAA \rightarrow Ggt$
919-E323 rev	TTTGACCCAAAACacCGGCGAGGCG	
919-D362 for	TGCCGGCGCAGTCGgtCGGCACTACA	$GAC \rightarrow Ggt$
919-D362 rev	TAATGTAGTGCCGacCGACTGCGCCG	
907-E117 for	TGATTGAGGTGGgtAGCGCGTTCCG	$GAA \rightarrow Ggt$
907-E117 rev	GGCGGAACGCGCTacCCACCTCAAT	

Underlined nucleotides code for glycine; the mutated nucleotides are in lower case.

To generate the 919-E255, 919-E323 and 919-E362 mutants, PCR was performed using 20ng of the pET 919-LOrf4 DNA as template, and the following primer pairs:

- 1) Orf4L for / 919-E255 rev
- 2) 919-E255 for / 919L rev

5

20

- 3) Orf4L for / 919-E323 rev
- 4) 919-E323 for / 919L rev
- 5) Orf4L for / 919-D362 rev
- 6) 919-D362 for / 919L rev

The second round of PCR was performed using the product of PCR 1-2, 3-4 or 5-6 as template, and as forward and reverse primers the "Orf4L for" and "919L rev" respectively.

For the mutant 907-E117, PCR have been performed using 200ng of chromosomal DNA of the 2996 strain as template and the following primer pairs:

- 7) 907L for / 907-E117 rev
- 15 8) 907-E117 for / 907L rev

The second round of PCR was performed using the products of PCR 7 and 8 as templates and the oligos "907L for" and "907L rev" as primers.

The PCR fragments containing each mutation were processed following the standard procedure, digested with *NdeI* and *XhoI* restriction enzymes and cloned into pET-21b+vector. The presence of each mutation was confirmed by sequence analysis.

Mutation of Glu117 to Gly in 907 is carried out similarly, as is mutation of residues Glu164, Ser213 and Asn348 in 922.

The E255G mutant of 919 shows a 50% reduction in activity; the E323G mutant shows a 70% reduction in activity; the E362G mutant shows no reduction in activity.

Example 4 – multimeric form

5

10

15

20

25

30

287-GST, 919^{untagged} and 953-His were subjected to gel filtration for analysis of quaternary structure or preparative purposes. The molecular weight of the native proteins was estimated using either FPLC Superose 12 (H/R 10/30) or Superdex 75 gel filtration columns (Pharmacia). The buffers used for chromatography for 287, 919 and 953 were 50 mM Tris-HCl (pH 8.0), 20 mM Bicine (pH 8.5) and 50 mM Bicine (pH 8.0), respectively.

Additionally each buffer contained 150-200 mM NaCl and 10% v/v glycerol. Proteins were dialysed against the appropriate buffer and applied in a volume of 200 μ l. Gel filtration was performed with a flow rate of 0.5 – 2.0 ml/min and the eluate monitored at 280nm. Fractions were collected and analysed by SDS-PAGE. Blue dextran 2000 and the molecular weight standards ribonuclease A, chymotrypsin A ovalbumin, albumin (Pharmacia) were used to calibrate the column. The molecular weight of the sample was estimated from a calibration curve of K_{av} vs. log M_{r} of the standards. Before gel filtration, 287-GST was digested with thrombin to cleave the GST moiety.

The estimated molecular weights for 287, 919 and 953-His were 73 kDa, 47 kDa and 43 kDa respectively. These results suggest 919 is monomeric while both 287 and 953 are principally dimeric in their nature. In the case of 953-His, two peaks were observed during gel filtration. The major peak (80%) represented a dimeric conformation of 953 while the minor peak (20%) had the expected size of a monomer. The monomeric form of 953 was found to have greater bactericidal activity than the dimer.

Example 5 - pSM214 and pET-24b vectors

953 protein with its native leader peptide and no fusion partners was expressed from the pET vector and also from pSM214 [Velati Bellini *et al.* (1991) *J. Biotechnol.* 18, 177-192].

The 953 sequence was cloned as a full-length gene into pSM214 using the *E. coli* MM294-1 strain as a host. To do this, the entire DNA sequence of the 953 gene (from ATG to the STOP codon) was amplified by PCR using the following primers:

953L for/2 CCGGAATTCTTATGAAAAAAATCATCTTCGCCGC Eco RI
953L rev/2 GCCCAAGCTTTTATTGTTTGGCTGCCTCGATT Hind III

which contain *Eco*RI and *Hind*III restriction sites, respectively. The amplified fragment was digested with *Eco*RI and *Hind*III and ligated with the pSM214 vector digested with the same two enzymes. The ligated plasmid was transformed into *E.coli* MM294-1 cells (by incubation in ice for 65 minutes at 37° C) and bacterial cells plated on LB agar containing 20µg/ml of chloramphenicol.

Recombinant colonies were grown over-night at 37°C in 4 ml of LB broth containing 20 µg/ml of chloramphenicol; bacterial cells were centrifuged and plasmid DNA extracted as and analysed by restriction with *Eco*RI and *Hind*III. To analyse the ability of the recombinant colonies to express the protein, they were inoculated in LB broth containing 20µg/ml of chloramphenicol and let to grown for 16 hours at 37°C. Bacterial cells were centrifuged and resuspended in PBS. Expression of the protein was analysed by SDS-PAGE and Coomassie Blue staining.

Expression levels were unexpectedly high from the pSM214 plasmid.

5

10

Oligos used to clone sequences into pSM-214 vectors were as follows:

ΔG287	Fwd	CCG <u>GAATTC</u> TTATG-TCGCCCGATGTTAAATCGGCGGA	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ2 287	Fwd	CCG <u>GAATTC</u> TTATG-AGCCAAGATATGGCGGCAGT	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TC A ATCCTGCTCTTTTTTGCCG	HindIII
Δ3 287	Fwd	CCG <u>GAATTC</u> TTATG-TCCGCCGAATCCGCAAATCA	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TC A ATCCTGCTCTTTTTTGCCG	HindIII
Δ4 287	Fwd	CCG <u>GAATTC</u> TTATG-GGAAGGGTTGATTTGGCTAATG	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TC A ATCCTGCTCTTTTTTGCCG	HindIII
Orf46.1	Fwd	CCG <u>GAATTC</u> TTATG-TCAGATTTGGCAAACGATTCTT	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTACGTATCATATTTCACGTGCTTC	HindIII
ΔG287-Orf46.1	Fwd	CCG <u>GAATTC</u> TTATG-TCGCCCGATGTTAAATCGGCGGA	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTACGTATCATATTTCACGTGCTTC	HindIII
919	Fwd	CCG <u>GAATTC</u> TTATG-CAAAGCAAGAGCATCCAAACCT	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTACGGGCGGTATTCGGGCT	HindIII
961L	Fwd	CCG <u>GAATTC</u> ATATG-AAACACTTTCCATCC	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTACCACTCGTAATTGAC	HindIII
961	Fwd	CCG <u>GAATTC</u> ATATG-GCCACAAGCGACGAC	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTACCACTCGTAATTGAC	HindIII
961c L	Fwd	CCG <u>GAATTC</u> TTATG-AAACACTTTCCATCC	EcoRI
pSM-214	Rev	GCCC <u>AAGCTT</u> -TCAACCCACGTTGTAAGGTTG	HindⅢ
961c	Fwd	CCG <u>GAATTC</u> TTATG-GCCACAAACGACGACG	EcoRI
pSM-214	Rev	GCCC <u>AAGCTT</u> -TCAACCCACGTTGTAAGGTTG	HindIII
953	Fwd	CCG <u>GAATTC</u> TTATG-GCCACCTACAAAGTGGACGA	EcoRI
(pSM-214)	Rev	GCCC <u>AAGCTT</u> -TTATTGTTTGGCTGCCTCGATT	HindIII

These sequences were manipulated, cloned and expressed as described for 953L.

For the pET-24 vector, sequences were cloned and the proteins expressed in pET-24 as described below for pET21. pET2 has the same sequence as pET-21, but with the kanamycin resistance cassette instead of ampicillin cassette.

5 Oligonucleotides used to clone sequences into pET-24b vector were:

ΔG 287 K	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC §	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC *	XhoI
Δ2 287 K	Fwd	CGCGGATCC <u>GCTAGC</u> -CAAGATATGGCGGCAGT §	NheI
Δ3 287 K	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA §	NheI
Δ4 287 Κ	Fwd	CGC <u>GCTAGC</u> -GGAAGGGTTGATTTGGCTAATGG §	NheI
Orf46.1 K	Fwd	GGGAATTC <u>CATATG</u> -GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCG <u>CTCGAG</u> - TTA CGTATCATATTTCACGTGC	XhoI
Orf46A K	Fwd	GGGAATTC <u>CATATG</u> -GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTCTATGCCTTGTGCGGCAT	XhoI
961 K	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAGCGACGACGA	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961a K	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCATTTAGCAATATTATCTTTGTTC	XhoI
961b K	Fwd	CGCGGATCC <u>CATATG</u> -AAAGCAAACAGTGCCGAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961c K	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCCG <u>CTCGAG</u> - TTA ACCCACGTTGTAAGGT	XhoI
961cL K	Fwd	CGCGGATCC <u>CATATG</u> -ATGAAACACTTTCCATCC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTAACCCACGTTGTAAGGT	XhoI
961d K	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCAGTCTGACACTGTTTTATCC	XhoI
ΔG 287-	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
919 K	Rev	CCCG <u>CTCGAG</u> -TTACGGGCGGTATTCGG	XhoI
ΔG 287-	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC	NheI
Orf46.1 K	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
ΔG 287-	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC	NheI
961 K	Rev	CCCG <u>CTCGAG</u> -TTACCACTCGTAATTGAC	XhoI

^{*} This primer was used as a Reverse primer for all the 287 forms.

Example 6 - ORF1 and its leader peptide

ORF1 from *N.meningitidis* (serogroup B, strain MC58) is predicted to be an outer membrane or secreted protein. It has the following sequence:

 $^{^{\}S}$ Forward primers used in combination with the $\Delta G278$ K reverse primer.

¹ MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN

-23-

	F-1		************			
	51			IEVYNKKGEL		
	101			NVDFGAEGRN		
	151			AEPVEMTSYM		
-	201	~		ASAYSWLVGG	~	
5	251			MFIYDAQKQK	~	
	301	QLVRKDWFYD	EIFAGDTHSV	FYEPRQNGKY	SFNDDNNGTG	KINAKHEHNS
	351	LPNRLKTRTV	QLFNVSLSET	AREPVYHAAG	GVNSYRPRLN	NGENISFIDE
	401	GKGELILTSN	INQGAGGLYF	QGDFTVSPEN	NETWQGAGVH	ISEDSTVTWK
	451	VNGVANDRLS	KIGKGTLHVQ	AKGENQGSIS	VGDGTVILDQ	QADDKGKKQA
10	501	FSEIGLVSGR	GTVQLNADNQ	FNPDKLYFGF	RGGRLDLNGH	SLSFHRIQNT
	551	DEGAMIVNHN	QDKESTVTIT	GNKDIATTGN	NNSLDSKKEI	AYNGWFGEKD
	601	TTKTNGRLNL	VYQPAAEDRT	LLLSGGTNLN	GNITQTNGKL	FFSGRPTPHA
	651	YNHLNDHWSQ	KEGIPRGEIV	WDNDWINRTF	KAENFQIKGG	QAVVSRNVAK
	701	VKGDWHLSNH	AQAVFGVAPH	QSHTICTRSD	WTGLTNCVEK	TITDDKVIAS
15	751	LTKTDISGNV	DLADHAHLNL	TGLATLNGNL	SANGDTRYTV	SHNATQNGNL
	801	SLVGNAQATF	NQATLNGNTS	ASGNASFNLS	DHAVQNGSLT	LSGNAKANVS
	851	HSALNGNVSL	ADKAVFHFES	SRFTGQISGG	KDTALHLKDS	EWTLPSGTEL
	901	GNLNLDNATI	TLNSAYRHDA	AGAQTGSATD	APRRRSRRSR	RSLLSVTPPT
	951	SVESRFNTLT	VNGKLNGQGT	FRFMSELFGY	RSDKLKLAES	SEGTYTLAVN
20	1001	NTGNEPASLE	QLTVVEGKDN	KPLSENLNFT	LQNEHVDAGA	WRYQLIRKDG
	1051	EFRLHNPVKE	QELSDKLGKA	EAKKQAEKDN	AQSLDALIAA	GRDAVEKTES
	1101	VAEPARQAGG	ENVGIMQAEE	EKKRVQADKD	TALAKQREAE	TRPATTAFPR
	1151	ARRARRDLPQ	LQPQPQPQPQ	RDLISRYANS	GLSEFSATLN	SVFAVQDELD
	1201	RVFAEDRRNA	VWTSGIRDTK	HYRSQDFRAY	RQQTDLRQIG	MQKNLGSGRV
25	1251	GILFSHNRTE	NTFDDGIGNS	ARLAHGAVFG	OYGIDRFYIG	ISAGAGFSSG
	1301	SLSDGIGGKI	RRRVLHYGIO	ARYRAGFGGF	GIEPHIGATR	YFVOKADYRY
	1351		~	ADYSFKPAOH		~
	1401			GVNAEIKGFT		
	1451	IKLGYRW*				~

30 The leader peptide is underlined.

A polymorphic form of ORF1 is disclosed in WO99/55873.

Three expression strategies have been used for ORF1:

- 1) ORF1 using a His tag, following WO99/24578 (ORF1-His);
- 2) ORF1 with its own leader peptide but without any fusion partner ('ORF1L'); and
- 35 3) ORF1 with the leader peptide (MKKTAIAIAVALAGFATVAQA) from *E.coli* OmpA ('Orf1LOmpA'):

MKKTAIAIAVALAGFATVAQAASAGHTYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSV VSRNGVAALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAE PVEMTSYMDGRKYIDONNYPDRVRIGAGROYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAONGSGGGTVNLGSEK 40 IKHSPYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQ NGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKG ELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGTLHVQAKGENQGSIS VGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIV NHNQDKESTVTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNGNIT 45 QTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLSNHA QAVFGVAPHQSHTICTRSDWTGLTNCVEKTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNLSANGDTRY TVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNGNVSLADKAV ${\tt FHFESSRFTGQISGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRS}$ LLSVTPPTSVESRFNTLTVNGKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKD 50 NKPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAOSLDALIAAGRDAVE KTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATTAFPRARRARRDLPQLQPQPQPQPQPDL ISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRV GILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQARYRAGF GGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAASGKVRTRVN 55 TAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLEAQHSAGIKLGYRW*

-24-

To make this construct, the clone pET911LOmpA (see below) was digested with the *Nhe*I and *Xho*I restriction enzymes and the fragment corresponding to the vector carrying the OmpA leader sequence was purified (pETLOmpA). The ORF1 gene coding for the mature protein was amplified using the oligonucleotides ORF1-For and ORF1-Rev (including the *Nhe*I and *Xho*I restriction sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOmpA fragment (see Figure 1). An additional AS dipeptide was introduced by the *Nhe*I site.

All three forms of the protein were expressed. The His-tagged protein could be purified and was confirmed as surface exposed, and possibly secreted (see Figure 3). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay.

ORF1LOmpA was purified as total membranes, and was localised in both the inner and outer membranes. Unexpectedly, sera raised against ORF1LOmpA show even better ELISA and anti-bactericidal properties than those raised against the His-tagged protein.

ORF1L was purified as outer membranes, where it is localised.

15 Example 7 – protein 911 and its leader peptide

Protein 911 from N.meningitidis (serogroup B, strain MC58) has the following sequence:

- 1 MKKNILEFWV GLFVLIGAAA VAFLAFRVAG GAAFGGSDKT YAVYADFGDI
- 51 GGLKVNAPVK SAGVLVGRVG AIGLDPKSYQ ARVRLDLDGK YQFSSDVSAQ
- 101 ILTSGLLGEQ YIGLQQGGDT ENLAAGDTIS VTSSAMVLEN LIGKFMTSFA
- 151 EKNADGGNAE KAAE*

The leader peptide is underlined.

5

10

20

Three expression strategies have been used for 911:

- 1) 911 with its own leader peptide but without any fusion partner ('911L');
- 2) 911 with the leader peptide from *E.coli* OmpA ('911LOmpA').
- To make this construct, the entire sequence encoding the OmpA leader peptide was included in the 5'- primer as a tail (primer 911LOmpA Forward). A *NheI* restriction site was inserted between the sequence coding for the OmpA leader peptide and the 911 gene encoding the predicted mature protein (insertion of one amino acid, a serine), to allow the use of this construct to clone different genes downstream the OmpA leader peptide sequence.
 - 3) 911 with the leader peptide (MKYLLPTAAAGLLLAAQPAMA) from *Erwinia carotovora* PelB ('911LpelB').

-25-

To make this construct, the 5'-end PCR primer was designed downstream from the leader sequence and included the *Nco*I restriction site in order to have the 911 fused directly to the PelB leader sequence; the 3'- end primer included the STOP codon. The expression vector used was pET22b+ (Novagen), which carries the coding sequence for the PelB leader peptide. The *Nco*I site introduces an additional methionine after the PelB sequence.

All three forms of the protein were expressed. ELISA titres were highest using 911L, with 919LOmpA also giving good results.

Example 8 - ORF46

5

25

10 The complete ORF46 protein from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```
LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRQ HFEPDGKYHL
                    FGSRGELAER SGHIGLGKIQ SHQLGNLMIQ QAAIKGNIGY IVRFSDHGHE
               101 VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
15
               151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLADRFH NAGSMLTQGV
               201
                    GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
               251
                    ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
               301 NAAQGIEAVS NIFMAAIPIK GIGAVRGKYG LGGITAHPIK RSQMGAIALP
               351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
20
               401
                    KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
               451
                    VSDAKPRWEV DRKLNKLTTR EQVEKNVQEI RNGNKNSNFS QHAQLEREIN
               501 KLKSADEINF ADGMGKFTDS MNDKAFSRLV KSVKENGFTN PVVEYVEING
               551 KAYIVRGNNR VFAAEYLGRI HELKFKKVDF PVPNTSWKNP TDVLNESGNV
               601 KRPRYRSK*
```

The leader peptide is underlined.

The sequences of ORF46 from other strains can be found in WO00/66741.

Three expression strategies have been used for ORF46:

- 1) ORF46 with its own leader peptide but without any fusion partner ('ORF46-2L');
- 30 2) ORF46 without its leader peptide and without any fusion partner ('ORF46-2'), with the leader peptide omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence:

```
SDLANDSFIR QVLDRQHFEP DGKYHLFGSR GELAERSGHI GLGKIQSHQL
                51
                    GNLMIQQAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS DEAGSPVDGF
35
               101
                    SLYRIHWDGY EHHPADGYDG PQGGGYPAPK GARDIYSYDI KGVAQNIRLN
                    LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPEL DRSGNAAEAF
               151
               201 NGTADIVKNI IGAAGEIVGA GDAVQGISEG SNIAVMHGLG LLSTENKMAR
                    INDLADMAQL KDYAAAAIRD WAVQNPNAAQ GIEAVSNIFM AAIPIKGIGA
               251
               301
                    VRGKYGLGGI TAHPIKRSQM GAIALPKGKS AVSDNFADAA YAKYPSPYHS
40
               351
                    RNIRSNLEQR YGKENITSST VPPSNGKNVK LADQRHPKTG VPFDGKGFPN
               401 FEKHVKYDTK LDIQELSGGG IPKAKPVSDA KPRWEVDRKL NKLTTREQVE
               451 KNVQEIRNGN KNSNFSQHAQ LEREINKLKS ADEINFADGM GKFTDSMNDK
               501 AFSRLVKSVK ENGFTNPVVE YVEINGKAYI VRGNNRVFAA EYLGRIHELK
               551 FKKVDFPVPN TSWKNPTDVL NESGNVKRPR YRSK*
```

3) ORF46 as a truncated protein, consisting of the first 433 amino acids ('ORF46.1L'), constructed by designing PCR primers to amplify a partial sequence corresponding to aa 1-433.

A STOP codon was included in the 3'-end primer sequences.

5

15

ORF46-2L is expressed at a very low level to *E.coli*. Removal of its leader peptide (ORF46-2) does not solve this problem. The truncated ORF46.1L form (first 433 amino acids, which are well conserved between serogroups and species), however, is well-expressed and gives excellent results in ELISA test and in the bactericidal assay.

ORF46.1 has also been used as the basis of hybrid proteins. It has been fused with 287, 919, and ORF1. The hybrid proteins were generally insoluble, but gave some good ELISA and bactericidal results (against the homologous 2996 strain):

Protein	ELISA	Bactericidal Ab
Orf1-Orf46.1-His	850	256
919-Orf46.1-His	12900	512
919-287-Orf46-His	n.d.	n.d.
Orf46.1-287His	150	8192
Orf46.1-919His	2800	2048
Orf46.1-287-919His	3200	16384

For comparison, 'triple' hybrids of ORF46.1, 287 (either as a GST fusion, or in Δ G287 form) and 919 were constructed and tested against various strains (including the homologous 2996 strain) *versus* a simple mixture of the three antigens. FCA was used as adjuvant:

	2996	BZ232	MC58	NGH38	F6124	BZ133
Mixture	8192	256	512	1024	>2048	>2048
ORF46.1-287-919his	16384	256	4096	8192	8192	8192
ΔG287-919-ORF46.1his	8192	64	4096	8192	8192	16384
ΔG287-ORF46.1-919his	4096	128	256	8192	512	1024

Again, the hybrids show equivalent or superior immunological activity.

Hybrids of two proteins (strain 2996) were compared to the individual proteins against various heterologous strains:

-27-

	1000	MC58	F6124 (MenA)
ORF46.1-His	<4	4096	<4
ORF1-His	8	256	128
ORF1—ORF46.1-His	1024	512	1024

Again, the hybrid shows equivalent or superior immunological activity.

Example 9 – protein 961

The complete 961 protein from *N.meningitidis* (serogroup B, strain MC58) has the following sequence:

5	1	MSMKHFPAKV	LTTAILATFC	SGALAATSDD	DVKKAATVAI	VAAYNNGQEI
	51	NGFKAGETIY	DIGEDGTITQ	KDATAADVEA	DDFKGLGLKK	VVTNLTKTVN
	101	ENKQNVDAKV	KAAESEIEKL	TTKLADTDAA	LADTDAALDE	TTNALNKLGE
	151	NITTFAEETK	TNIVKIDEKL	EAVADTVDKH	AEAFNDIADS	LDETNTKADE
	201	AVKTANEAKQ	TAEETKQNVD	AKVKAAETAA	GKAEAAAGTA	NTAADKAEAV
10	251	AAKVTDIKAD	IATNKADIAK	NSARIDSLDK	NVANLRKETR	QGLAEQAALS
	301	GLFQPYNVGR	FNVTAAVGGY	KSESAVAIGT	GFRFTENFAA	KAGVAVGTSS
	351	GSSAAYHVGV	NYEW*			

The leader peptide is underlined.

- 15 Three approaches to 961 expression were used:
 - 1) 961 using a GST fusion, following WO99/57280 ('GST961');
 - 2) 961 with its own leader peptide but without any fusion partner ('961L'); and
 - 3) 961 without its leader peptide and without any fusion partner ('961^{untagged}'), with the leader peptide omitted by designing the 5'-end PCR primer downstream from the predicted leader sequence.

All three forms of the protein were expressed. The GST-fusion protein could be purified and antibodies against it confirmed that 961 is surface exposed (Figure 4). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay. 961L could also be purified and gave very high ELISA titres.

25 Protein 961 appears to be phase variable. Furthermore, it is not found in all strains of *N.meningitidis*.

Example 10 – protein 287

20

Protein 287 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

30 EFERSVIAMA CIFALSACGG GGGSPDVKS ADTLSKPAAP VVAEKETEVK
51 EDAPQAGSQG QGAPSTQGSQ DMAAVSAENT GNGGAATTDK PKNEDEGPQN
101 DMPQNSAESA NQTGNNQPAD SSDSAPASNP APANGGSNFG RVDLANGVLI
151 DGPSQNITLT HCKGDSCNGD NLLDEEAPSK SEFENLNESE RIEKYKKDGK

-28-

```
201 SDKFTNLVAT AVQANGTNKY VIIYKDKSAS SSSARFRRSA RSRRSLPAEM
251 PLIPVNQADT LIVDGEAVSL TGHSGNIFAP EGNYRYLTYG AEKLPGGSYA
301 LRVQGEPAKG EMLAGTAVYN GEVLHFHTEN GRPYPTRGRF AAKVDFGSKS
351 VDGIIDSGDD LHMGTQKFKA AIDGNGFKGT WTENGGGDVS GRFYGPAGEE
401 VAGKYSYRPT DAEKGGFGVF AGKKEQD*
```

The leader peptide is shown underlined.

5

30

The sequences of 287 from other strains can be found in Figures 5 and 15 of WO00/66741.

Example 9 of WO99/57280 discloses the expression of 287 as a GST-fusion in E.coli.

- 10 A number of further approaches to expressing 287 in *E.coli* have been used, including:
 - 1) 287 as a His-tagged fusion ('287-His');
 - 2) 287 with its own leader peptide but without any fusion partner ('287L');
 - 3) 287 with the ORF4 leader peptide and without any fusion partner ('287LOrf4'); and
 - 4) 287 without its leader peptide and without any fusion partner ('287^{untagged}'):

```
1 CGGGGGGSPD VKSADTLSKP AAPVVAEKET EVKEDAPQAG SQGQGAPSTQ
51 GSQDMAAVSA ENTGNGGAAT TDKPKNEDEG PQNDMPQNSA ESANQTGNNQ
101 PADSSDSAPA SNPAPANGGS NFGRVDLANG VLIDGPSQNI TLTHCKGDSC
151 NGDNLLDEEA PSKSEFENLN ESERIEKYKK DGKSDKFTNL VATAVQANGT
201 NKYVIIYKDK SASSSARFR RSARSRSLP AEMPLIPVNQ ADTLIVDGEA
251 VSLTGHSGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP AKGEMLAGTA
301 VYNGEVLHFH TENGRPYPTR GRFAAKVDFG SKSVDGIIDS GDDLHMGTQK
351 FKAAIDGNGF KGTWTENGGG DVSGRFYGPA GEEVAGKYSY RPTDAEKGGF
```

25 All these proteins could be expressed and purified.

'287L' and '287LOrf4' were confirmed as lipoproteins.

As shown in Figure 2, '287LOrf4' was constructed by digesting 919LOrf4 with *Nhe*I and *Xho*I. The entire ORF4 leader peptide was restored by the addition of a DNA sequence coding for the missing amino acids, as a tail, in the 5'-end primer (287LOrf4 for), fused to 287 coding sequence. The 287 gene coding for the mature protein was amplified using the oligonucleotides 287LOrf4 For and Rev (including the *Nhe*I and *Xho*I sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOrf4 fragment.

Example 11 - further non-fusion proteins with/without native leader peptides

A similar approach was adopted for *E.coli* expression of further proteins from WO99/24578, WO99/36544 and WO99/57280.

WO 01/64922

5

20

25

The following were expressed without a fusion partner: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 982, and Orf143-1. Protein 117-1 was confirmed as surface-exposed by FACS and gave high ELISA titres.

The following were expressed with the native leader peptide but without a fusion partner: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 926, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1. These proteins are given the suffix 'L'.

His-tagged protein 760 was expressed with and without its leader peptide. The deletion of the signal peptide greatly increased expression levels. The protein could be purified most easily using 2M urea for solubilisation.

His-tagged protein 264 was well-expressed using its own signal peptide, and the 30kDa protein gave positive Western blot results.

All proteins were successfully expressed.

The localisation of 593, 121-1, 128-1, 593, 726, and 982 in the cytoplasm was confirmed.

The localisation of 920-1L, 953L, ORF9-1L, ORF85-2L, ORF97-1L, 570L, 580L and 664L in the periplasm was confirmed.

The localisation of ORF40L in the outer membrane, and 008 and 519-1L in the inner membrane was confirmed. ORF25L, ORF4L, 406L, 576-1L were all confirmed as being localised in the membrane.

Protein 206 was found not to be a lipoprotein.

ORF25 and ORF40 expressed with their native leader peptides but without fusion partners, and protein 593 expressed without its native leader peptide and without a fusion partner, raised good anti-bactericidal sera. Surprisingly, the forms of ORF25 and ORF40 expressed without fusion partners and using their own leader peptides (*i.e.* 'ORF25L' and 'ORF40L') give better results in the bactericidal assay than the fusion proteins.

Proteins 920L and 953L were subjected to N-terminal sequencing, giving hrvwvetah and atykvdeyhanarfaf, respectively. This sequencing confirms that the predicted leader peptides were cleaved and, when combined with the periplasmic location, confirms that the

proteins are correctly processed and localised by *E.coli* when expressed from their native leader peptides.

The N-terminal sequence of protein 519.1L localised in the inner membrane was MEFFIILLA, indicating that the leader sequence is not cleaved. It may therefore function as both an uncleaved leader sequence and a transmembrane anchor in a manner similar to the leader peptide of PBP1 from *N.gonorrhoeae* [Ropp & Nicholas (1997) *J. Bact.* 179:2783-2787.]. Indeed the N-terminal region exhibits strong hydrophobic character and is predicted by the Tmpred. program to be transmembrane.

Example 12 - lipoproteins

5

10 The incorporation of palmitate in recombinant lipoproteins was demonstrated by the method of Kraft et. al. [J. Bact. (1998) 180:3441-3447.]. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100µg/ml) liquid culture. The culture was diluted to an OD₅₅₀ of 0.1 in 5.0 ml of fresh medium LB/Amp medium containing 5 µC/ml [³H] palmitate (Amersham). When the OD₅₅₀ of the culture reached 0.4-15 0.8, recombinant lipoprotein was induced for 1 hour with IPTG (final concentration 1.0 mM). Bacteria were harvested by centrifugation in a bench top centrifuge at 2700g for 15 min and washed twice with 1.0 ml cold PBS. Cells were resuspended in 120µl of 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1.0% w/v SDS and lysed by boiling for 10 min. After centrifugation at 13000g for 10 min the supernatant was collected and proteins precipitated 20 by the addition of 1.2 ml cold acetone and left for 1 hour at -20 °C. Protein was pelleted by centrifugation at 13000g for 10 min and resuspended in 20-50µl (calculated to standardise loading with respect to the final O.D of the culture) of 1.0% w/v SDS. An aliquot of 15 µl was boiled with 5µl of SDS-PAGE sample buffer and analysed by SDS-PAGE. After electrophoresis gels were fixed for 1 hour in 10% v/v acetic acid and soaked for 30 minutes 25 in Amplify solution (Amersham). The gel was vacuum-dried under heat and exposed to Hyperfilm (Kodak) overnight -80 °C.

Incorporation of the [³H] palmitate label, confirming lipidation, was found for the following proteins: Orf4L, Orf25L, 287L, 287LOrf4, 406.L, 576L, 926L, 919L and 919LOrf4.

Example 13 – domains in 287

Based on homology of different regions of 287 to proteins that belong to different functional classes, it was split into three 'domains', as shown in Figure 5. The second domain shows

WO 01/64922

5

15

homology to IgA proteases, and the third domain shows homology to transferrin-binding proteins.

Each of the three 'domains' shows a different degree of sequence conservation between *N.meningitidis* strains – domain C is 98% identical, domain A is 83% identical, whilst domain B is only 71% identical. Note that protein 287 in strain MC58 is 61 amino acids longer than that of strain 2996. An alignment of the two sequences is shown in Figure 7, and alignments for various strains are disclosed in WO00/66741 (see Figures 5 and 15 therein).

The three domains were expressed individually as C-terminal His-tagged proteins. This was done for the MC58 and 2996 strains, using the following constructs:

10 287a-MC58 (aa 1-202), 287b-MC58 (aa 203-288), 287c-MC58 (aa 311-488). 287a-2996 (aa 1-139), 287b-2996 (aa 140-225), 287c-2996 (aa 250-427).

To make these constructs, the stop codon sequence was omitted in the 3'-end primer sequence. The 5' primers included the *NheI* restriction site, and the 3' primers included a *XhoI* as a tail, in order to direct the cloning of each amplified fragment into the expression vector pET21b+ using *NdeI-XhoI*, *NheI-XhoI* or *NdeI-HindIII* restriction sites.

All six constructs could be expressed, but 287b-MC8 required denaturation and refolding for solubilisation.

Deletion of domain A is described below (' $\Delta 4$ 287-His').

Immunological data (serum bactericidal assay) were also obtained using the various domains 20 from strain 2996, against the homologous and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-His	32000	16	4096	4096	512	8000	16000
287(B)-His	256	-	-	-	-	16	-
287(C)-His	256	_	32	512	32	2048	>2048
287(B-C)-His	64000	128	4096	64000	1024	64000	32000

Using the domains of strain MC58, the following results were obtained: .

	MC58	2996	BZ232	NGH38	394/98	MenA	MenC
287-His	4096	32000	16	4096	512	8000	16000
287(B)-His	128	128	_	-	-	-	128
287(C)-His	_	16	-	1024	_	512	-
287(B-C)-His	16000	64000	128	64000	512	64000	>8000

Example 14 – deletions in 287

WO 01/64922

5

As well as expressing individual domains, 287 was also expressed (as a C-terminal His-tagged protein) by making progressive deletions within the first domain. These

Four deletion mutants of protein 287 from strain 2996 were used (Figure 6):

- 1) '287-His', consisting of amino acids 18-427 (i.e. leader peptide deleted);
- 2) ' Δ 1 287-His', consisting of amino acids 26-427;
- 3) ' Δ 2 287-His', consisting of amino acids 70-427;
- 4) ' Δ 3 287-His', consisting of amino acids 107-427; and
- 5) 'Δ4 287-His', consisting of amino acids 140-427 (=287-bc).
- 10 The ' Δ 4' protein was also made for strain MC58 (' Δ 4 287MC58-His'; aa 203-488).

The constructs were made in the same way as 287a/b/c, as described above.

All six constructs could be expressed and protein could be purified. Expression of 287-His was, however, quite poor.

Expression was also high when the C-terminal His-tags were omitted.

15 Immunological data (serum bactericidal assay) were also obtained using the deletion mutants, against the homologous (2996) and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-his	32000	16	4096	4096	512	8000	16000
Δ1 287-His	16000	128	4096	4096	1024	8000	16000
Δ2 287-His	16000	128	4096	>2048	512	16000	>8000
Δ3 287-His	16000	128	4096	>2048	512	16000	>8000
Δ4 287-His	64000	128	4096	64000	1024	64000	32000

The same high activity for the $\Delta 4$ deletion was seen using the sequence from strain MC58.

As well as showing superior expression characteristics, therefore, the mutants are immunologically equivalent or superior.

Example 15 - poly-glycine deletions

5

30

The ' $\Delta 1$ 287-His' construct of the previous example differs from 287-His and from ' 287^{untagged} ' only by a short N-terminal deletion (GGGGGGS). Using an expression vector which replaces the deleted serine with a codon present in the *Nhe* cloning site, however, this amounts to a deletion only of (Gly)₆. Thus, the deletion of this (Gly)₆ sequence has been shown to have a dramatic effect on protein expression.

The protein lacking the N-terminal amino acids up to GGGGGG is called ' Δ G 287'. In strain MC58, its sequence (leader peptide underlined) is:

			→ ΔG28	37	
1	MFKRSVIAMA	CIFALSACGG	GGGGSPDVKS	ADTLSKPAAP	VVSEKETEAK
51	EDAPQAGSQG	QGAPSAQGSQ	DMAAVSEENT	${\tt GNGGAVTADN}$	PKNEDEVAQN
101	DMPQNAAGTD	SSTPNHTPDP	NMLAGNMENQ	ATDAGESSQP	ANQPDMANAA
151	DGMQGDDPSA	GGQNAGNTAA	QGANQAGNNQ	AAGSSDPIPA	SNPAPANGGS
201	NFGRVDLANG	VLIDGPSQNI	TLTHCKGDSC	SGNNFLDEEV	QLKSEFEKLS
251	DADKISNYKK	DGKNDKFVGL	VADSVQMKGI	NQYIIFYKPK	PTSFARFRRS
301	ARSRRSLPAE	MPLIPVNQAD	TLIVDGEAVS	LTGHSGNIFA	PEGNYRYLTY
351	GAEKLPGGSY	ALRVQGEPAK	GEMLAGAAVY	NGEVLHFHTE	NGRPYPTRGR
401	FAAKVDFGSK	SVDGIIDSGD	DLHMGTQKFK	AAIDGNGFKG	TWTENGSGDV
451	SGKFYGPAGE	EVAGKYSYRP	TDAEKGGFGV	FAGKKEQD*	
	101 151 201 251 301 351 401	51 EDAPQAGSQG 101 DMPQNAAGTD 151 DGMQGDDPSA 201 NFGRVDLANG 251 DADKISNYKK 301 ARSRRSLPAE 351 GAEKLPGGSY 401 FAAKVDFGSK	51 EDAPQAGSQG QGAPSAQGSQ 101 DMPQNAAGTD SSTPNHTPDP 151 DGMQGDDPSA GGQNAGNTAA 201 NFGRVDLANG VLIDGPSQNI 251 DADKISNYKK DGKNDKFVGL 301 ARSRRSLPAE MPLIPVNQAD 351 GAEKLPGGSY ALRVQGEPAK 401 FAAKVDFGSK SVDGIIDSGD	1 MFKRSVIAMA CIFALSACGG GGGSPDVKS 51 EDAPQAGSQG QGAPSAQGSQ DMAAVSEENT 101 DMPQNAAGTD SSTPNHTPDP NMLAGNMENQ 151 DGMQGDDPSA GGQNAGNTAA QGANQAGNNQ 201 NFGRVDLANG VLIDGPSQNI TLTHCKGDSC 251 DADKISNYKK DGKNDKFVGL VADSVQMKGI 301 ARSRRSLPAE MPLIPVNQAD TLIVDGEAVS 351 GAEKLPGGSY ALRVQGEPAK GEMLAGAAVY 401 FAAKVDFGSK SVDGIIDSGD DLHMGTQKFK	101 DMPQNAAGTD SSTPNHTPDP NMLAGNMENQ ATDAGESSQP 151 DGMQGDDPSA GGQNAGNTAA QGANQAGNNQ AAGSSDPIPA 201 NFGRVDLANG VLIDGPSQNI TLTHCKGDSC SGNNFLDEEV 251 DADKISNYKK DGKNDKFVGL VADSVQMKGI NQYIIFYKPK 301 ARSRRSLPAE MPLIPVNQAD TLIVDGEAVS LTGHSGNIFA 351 GAEKLPGGSY ALRVQGEPAK GEMLAGAAVY NGEVLHFHTE 401 FAAKVDFGSK SVDGIIDSGD DLHMGTQKFK AAIDGNGFKG

 Δ G287, with or without His-tag (' Δ G287-His' and ' Δ G287K', respectively), are expressed at very good levels in comparison with the '287-His' or '287 ^{untagged}'.

On the basis of gene variability data, variants of $\Delta G287$ -His were expressed in *E.coli* from a number of MenB strains, in particular from strains 2996, MC58, 1000, and BZ232. The results were also good.

It was hypothesised that poly-Gly deletion might be a general strategy to improve expression. Other MenB lipoproteins containing similar (Gly)_n motifs (near the N-terminus, downstream of a cysteine) were therefore identified, namely Tbp2 (NMB0460), 741 (NMB 1870) and 983 (NMB1969):

```
TBP2
                                            → ΔGTbp2
                MNNPLVNQAA MVLPVFLLSA CLGGGGSFDL DSVDTEAPRP APKYQDVFSE
             1
                KPQAQKDQGG YGFAMRLKRR NWYPQAKEDE VKLDESDWEA TGLPDEPKEL
35
                PKRQKSVIEK VETDSDNNIY SSPYLKPSNH QNGNTGNGIN QPKNQAKDYE
           101
                NFKYVYSGWF YKHAKREFNL KVEPKSAKNG DDGYIFYHGK EPSROLPASG
           201
                KITYKGVWHF ATDTKKGQKF REIIQPSKSQ GDRYSGFSGD DGEEYSNKNK
                STLTDGQEGY GFTSNLEVDF HNKKLTGKLI RNNANTDNNQ ATTTQYYSLE
           251
           301
                AQVTGNRFNG KATATDKPQQ NSETKEHPFV SDSSSLSGGF FGPQGEELGF
40
           351 RFLSDDQKVA VVGSAKTKDK PANGNTAAAS GGTDAAASNG AAGTSSENGK
           401
                LTTVLDAVEL KLGDKEVQKL DNFSNAAQLV VDGIMIPLLP EASESGNNOA
           451 NQGTNGGTAF TRKFDHTPES DKKDAQAGTQ TNGAQTASNT AGDTNGKTKT
```

-34-

	501	YEVEVCCSNL NYLKYGMLTR KNSKSAMQAG ESSSQADAKT EQVEQSMFLQ
	551	GERTDEKEIP SEQNIVYRGS WYGYIANDKS TSWSGNASNA TSGNRAEFTV
	601	Z Z X
_	651	~
5	701	VVFGAKRQQP VR*
	741	→ ΔG741
	1	VNRTAFCCLS LTTALILTAC SSGGGGVAAD IGAGLADALT APLDHKDKGL
10	51	QSLTLDQSVR KNEKLKLAAQ GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ
10	101	IEVDGQLITL ESGEFQVYKQ SHSALTAFQT EQIQDSEHSG KMVAKRQFRI
	151	GDIAGEHTSF DKLPEGGRAT YRGTAFGSDD AGGKLTYTID FAAKQGNGKI
	201	EHLKSPELNV DLAAADIKPD GKRHAVISGS VLYNQAEKGS YSLGIFGGKA
	251	QEVAGSAEVK TVNGIRHIGL AAKQ*
15	983	0002
13	983	► AG983
	51	THE PERSON OF TH
	101	
	151	LYGRKEHGYN ENYKNYTAYM RKEAPEDGGG KDIEASFDDE AVIETEAKPT
20	201	DIRHVKEIGH IDLVSHIIGG RSVDGRPAGG IAPDATLHIM NTNDETKNEM
	251	MVAAIRNAWV KLGERGVRIV NNSFGTTSRA GTADLFQIAN SEEQYRQALL
	301	
	351	LLPFYEKDAQ KGIITVAGVD RSGEKFKREM YGEPGTEPLE YGSNHCGITA
	401	MWCLSAPYEA SVRFTRTNPI QIAGTSFSAP IVTGTAALLL QKYPWMSNDN
25	451	LRTTLLTTAQ DIGAVGVDSK FGWGLLDAGK AMNGPASFPF GDFTADTKGT
	501	
	551	KSDMRVETKG ALIYNGAASG GSLNSDGIVY LADTDQSGAN ETVHIKGSLQ
	601	LDGKGTLYTR LGKLLKVDGT AIIGGKLYMS ARGKGAGYLN STGRRVPFLS
	651	AAKIGQDYSF FTNIETDGGL LASLDSVEKT AGSEGDTLSY YVRRGNAART
30	701	ASAAAHSAPA GLKHAVEQGG SNLENLMVEL DASESSATPE TVETAAADRT
	751	DMPGIRPYGA TFRAAAAVQH ANAADGVRIF NSLAATVYAD STAAHADMQG
	801	RRLKAVSDGL DHNGTGLRVI AQTQQDGGTW EQGGVEGKMR GSTQTVGIAA
	851	KTGENTTAAA TLGMGRSTWS ENSANAKTDS ISLFAGIRHD AGDIGYLKGL
0.5	901	FSYGRYKNSI SRSTGADEHA EGSVNGTLMQ LGALGGVNVP FAATGDLTVE
35	951	GGLRYDLLKQ DAFAEKGSAL GWSGNSLTEG TLVGLAGLKL SQPLSDKAVL
	1001	
	1051	FGNGWNGLAR YSYAGSKQYG NHSGRVGVGY RF*

Tbp2 and 741 genes were from strain MC58; 983 and 287 genes were from strain 2996.

40 These were cloned in pET vector and expressed in *E.coli* without the sequence coding for their leader peptides or as "ΔG forms", both fused to a C-terminal His-tag. In each case, the same effect was seen – expression was good in the clones carrying the deletion of the poly-glycine stretch, and poor or absent if the glycines were present in the expressed protein:

ORF	Express.	Purification	Bact. Activity
287-His(2996)	+/-	+	+
'287 ^{untagged} ', (2996)	+/-	nd	nd
ΔG287-His(2996)	+	+	+
ΔG287K(2996)	+	+	+
Δ G287-His(MC58)	+	+	+
Δ G287-His(1000)	+	+	+
Δ G287-His(BZ232)	+	+	+
Tbp2-His(MC58)	+/-	nd	nd
Δ GTbp2-His(MC58)	+	+	
741-His(MC58)	+/-	nd	nd
Δ G741-His(MC58)	+	+	
983-His (2996)			
ΔG983-His (2996)	+	+	

SDS-PAGE of the proteins is shown in Figure 13.

△G287 and hybrids

5

 $\Delta G287$ proteins were made and purified for strains MC58, 1000 and BZ232. Each of these gave high ELISA titres and also serum bactericidal titres of >8192. $\Delta G287K$, expressed from pET-24b, gave excellent titres in ELISA and the serum bactericidal assay. $\Delta G287$ -ORF46.1K may also be expressed in pET-24b.

 Δ G287 was also fused directly in-frame upstream of 919, 953, 961 (sequences shown below) and ORF46.1:

	ΔG287-9	19				
10	1	ATGGCTAGCC	CCGATGTTAA	ATCGGCGGAC	ACGCTGTCAA	AACCGGCCGC
	51	TCCTGTTGTT	GCTGAAAAAG	AGACAGAGGT	AAAAGAAGAT	GCGCCACAGG
	101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCA	CACAAGGCAG	CCAAGATATG
	151	GCGGCAGTTT	CGGCAGAAAA	TACAGGCAAT	GGCGGTGCGG	CAACAACGGA
	201	CAAACCCAAA	AATGAAGACG	AGGGACCGCA	AAATGATATG	CCGCAAAATT
15	251	CCGCCGAATC	CGCAAATCAA	ACAGGGAACA	ACCAACCCGC	CGATTCTTCA
	301	GATTCCGCCC	CCGCGTCAAA	CCCTGCACCT	GCGAATGGCG	GTAGCAATTT
	351	TGGAAGGGTT	${\tt GATTTGGCTA}$	ATGGCGTTTT	GATTGATGGG	CCGTCGCAAA
	401	ATATAACGTT	GACCCACTGT	AAAGGCGATT	CTTGTAATGG	TGATAATTTA
20	451	TTGGATGAAG	AAGCACCGTC	AAAATCAGAA	TTTGAAAATT	TAAATGAGTC
20	501	TGAACGAATT	GAGAAATATA	AGAAAGATGG	GAAAAGCGAT	AAATTTACTA
	551	ATTTGGTTGC	GACAGCAGTT	CAAGCTAATG	${\tt GAACTAACAA}$	ATATGTCATC
	601	ATTTATAAAG	ACAAGTCCGC	TTCATCTTCA	TCTGCGCGAT	TCAGGCGTTC
	651	TGCACGGTCG	AGGAGGTCGC	TTCCTGCCGA	${\tt GATGCCGCTA}$	ATCCCCGTCA
0.5	701	ATCAGGCGGA	TACGCTGATT	GTCGATGGGG	AAGCGGTCAG	CCTGACGGGG
25	751	CATTCCGGCA	ATATCTTCGC	GCCCGAAGGG	AATTACCGGT	ATCTGACTTA
	801	CGGGGCGGAA	AAATTGCCCG	GCGGATCGTA	TGCCCTCCGT	GTGCAAGGCG
	851	AACCGGCAAA	AGGCGAAATG	CTTGCTGGCA	CGGCCGTGTA	CAACGGCGAA
	901	GTGCTGCATT	$\mathtt{TTCATACGGA}$	AAACGGCCGT	CCGTACCCGA	CTAGAGGCAG
20	951	GTTTGCCGCA	AAAGTCGATT	TCGGCAGCAA	ATCTGTGGAC	GGCATTATCG
30	1001	ACAGCGGCGA	TGATTTGCAT	ATGGGTACGC	AAAAATTCAA	AGCCGCCATC

-36-

	1051	CATCCAAACC	$CCTTTT \lambda \lambda CCC$	GACTTGGACG	CANAMECCC	CCCCCCATCT
	1101			CGGCCGGCGA		
	1151			GAAAAGGGCG		
	1201			CGGAGGAGGA		
5	1251			ACACATCCGT		
	1301	CGGTCGGCAT	CCCCGACCCC	GCCGGAACGA	CGGTCGGCGG	CGGCGGGGCC
	1351			CCTGTCCCTG		
	1401			CCTTCCGCCT		
10	1451			GTGTGCGCCC		
10	1501			GTTTTTTGAA		
	1551			TTGCCGGTAC		
	1601			AGGCGGACGG		
	1651 1701			TATCTCCGTC GCATCAGGCA		
15	1751			ACACATACCG		
10	1801			AATCAAAGGC		
	1851			ACCAAATCAA		
	1901			GCCGAAGACC		
	1951			TCTGAAAACC		
20	2001	CATCGGCTAT	GCCGACAAAA	ACGAACATCC	CTACGTTTCC	ATCGGACGCT
	2051	ATATGGCGGA	CAAAGGCTAC	CTCAAGCTCG	GGCAGACCTC	GATGCAGGGC
	2101	ATCAAAGCCT	ATATGCGGCA	AAATCCGCAA	CGCCTCGCCG	AAGTTTTGGG
	2151			TTTTCCGCGA		
05	2201			GGCACGCCGT		
25	2251			TACCTTGGGC		
	2301			AAGCCCTCAA		
	2351			GGCGCGGTGC ACTTGCCGGC		
	2401 2451			ACGGTATGAA		
30	2501	TCGAG	CICCIACCA	ACGGIAIGAA	GCCCGAATAC	CGCCCGIAAC
	2302	1.00110				
	1	MASPDVKSAD	TLSKPAAPVV	AEKETEVKED	APQAGSQGQG	APSTQGSQDM
	51	AAVSAENTGN	${\tt GGAATTDKPK}$	NEDEGPQNDM	PQNSAESANQ	TGNNQPADSS
0.5	101			DLANGVLIDG		
35	151			EKYKKDGKSD		
	201			RRSLPAEMPL		
	251			KLPGGSYALR		
	301			KVDFGSKSVD		
40	351 401			FYGPAGEEVA PQPDTSVING		
40	451			SLOSFRLGCA		
	501			NGSLAGTVTG	~ ~	~ ~
	551			ALVRIROTGK		
	601			HTRNQINGGA		
45	651			ADKNEHPYVS		
	701	IKAYMRQNPQ	RLAEVLGQNP	SYIFFRELAG	SSNDGPVGAL	GTPLMGEYAG
	751	AVDRHYITLG	APLFVATAHP	VTRKALNRLI	MAQDTGSAIK	GAVRVDYFWG
	801	YGDEAGELAG	KQKTTGYVWQ	LLPNGMKPEY	RP*	
50						
30	∆G287-9	53				
	1		СССАТСТТАА	ATCGGCGGAC	ACGCTGTCAA	AACCGGCCGC
	51			AGACAGAGGT		
	101			GCGCCATCCA		
55	151			TACAGGCAAT		
	201			AGGGACCGCA		
	251	CCGCCGAATC	CGCAAATCAA	ACAGGGAACA	ACCAACCCGC	CGATTCTTCA
	301			CCCTGCACCT		
(0	351			ATGGCGTTTT		
60	401			AAAGGCGATT		
	451			AAAATCAGAA		
	501			AGAAAGATGG		
	551 601			CAAGCTAATG		
65	601 651			TTCATCTTCA TTCCTGCCGA		
00	701			GTCGATGGGG		
	751			GCCCGAAGGG		
	, , , ,	JIII I CCOOCA	111111011060	CCCCOARGGG	THE THICKS!	AICIGACIIA

-37-

	801	CGGGGGGGAA	<u>እ እ አጥጥር</u> ርርርር	CCCC A TCCTA	TGCCCTCCGT	GTGC A AGGCG
	851				CGGCCGTGTA	
	901				CCGTACCCGA	
	951				ATCTGTGGAC	
5	1001				AAAAATTCAA	
	1051	GATGGAAACG	GCTTTAAGGG	GACTTGGACG	GAAAATGGCG	GCGGGGATGT
	1101	TTCCGGAAGG	TTTTACGGCC	CGGCCGGCGA	GGAAGTGGCG	GGAAAATACA
	1151				GATTCGGCGT	
10	1201				GGAGCCACCT	
10	1251				CGACCATTTC	
	1301 1351				GTTCCGTCGA	
	1401				ATCCCCGTTG GAAATCAGCC	
	1451				TTTCCACCAA	
15	1501				AACCTGACCA	
10	1551				ATTCAACTGC	
	1601				ACTTCAGCAC	
	1651	CGCACCAAAT	GGGGCGTGGA	CTACCTCGTT	AACGTTGGTA	TGACCAAAAG
	1701	CGTCCGCATC	GACATCCAAA	TCGAGGCAGC	CAAACAATAA	CTCGAG
20						
	1.				APQAGSQGQG	
	51				PQNSAESANQ	-
	101				PSQNITLTHC	
25	151				KFTNLVATAV	
23	201 251				IPVNQADTLI VQGEPAKGEM	
	301				GIIDSGDDLH	
	351				GKYSYRPTDA	
	401				NTSTNVGGFY	
30	451				DIFDAAQYPD	~
	501				YQSPMAKTEV	
	551	RTKWGVDYLV	NVGMTKSVRI	DIQIEAAKQ*		
35	∧G287—9 <i>6</i>	51.				
35	<u>ΔG287-96</u>		ССGАТGТТАА	ATCGGCGGAC	ACGCTGTCAA	AACCGGCCGC
35	<u>ΔG287−96</u> 1 51	ATGGCTAGCC			ACGCTGTCAA AAAAGAAGAT	
35	1	ATGGCTAGCC TCCTGTTGTT	GCTGAAAAAG	AGACAGAGGT	ACGCTGTCAA AAAAGAAGAT CACAAGGCAG	GCGCCACAGG
	1 51	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA	GCTGAAAAAG AGGACAGGGC	AGACAGAGGT GCGCCATCCA	AAAAGAAGAT	GCGCCACAGG CCAAGATATG
35 40	1 51 101	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT
	1 51 101 151 201 251	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA
	1 51 101 151 201 251 301	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC	GCTGAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCGG GCGAATGGCG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT
	1 51 101 151 201 251 301 351	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA
40	1 51 101 151 201 251 301 351 401	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA
	1 51 101 151 201 251 301 351 401 451	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC
40	1 51 101 151 201 251 301 351 401 451 501	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA
40	1 51 101 151 201 251 301 351 401 451 501	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC
40	1 51 101 151 201 251 301 351 401 451 501	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCATCTTCA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC
40	1 51 101 151 201 251 301 351 401 451 501 551 601	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTTATAAAG TGCACGGTCG	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCATCTTCA TTCCTGCCGA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 651	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCCGA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATG CAAGCTAATG TTCATCTTCA TTCCTGCCGA GTCGATGGGG	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG TTTGAAAATT GAAAAGCGAT GAACAAACCAA TCTGCGCGAT GATGCCGCTA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCCGA CATTCCGGCA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAAGCTAATG TTCATCTTCA TTCCTGCCGA GTCGATGGGG GCCCGAAGGG	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATGCCGCTA AAGCGGTCAG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 851	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGGA CATTCCGGCA CGGGGCGGAA AACCGGCAAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA AAGCACCGTC GAGAAATATA GACAGCAGTT ACAGCAGTT ACAGCAGTT ACAGCAGTT ACAGCAGTT ACAGTCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCCTGCCGA GTCGATGGGG GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGCCA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATGCGCGTA AAGCGGTCAG AATTACCGGT TGCCCCTCCGT CGGCCGTGTA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCCGA CATTCCGGCA CGGGGCGGAA AACCGGCAAA GTGCTGCATT	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA AACCACCTGT AAGCACCGTC GAGAAATAA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAAGCTAATG TTCATCTTCA TTCCTGCCGA GTCGATGGG GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGGCA AAACGGCCGT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATGCGCTA AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAAT ATTTGGTTGC ATTTATAAAG TGCACGGCGA CATTCCGGCA CGGGGCGGAA AACCGGCAAA GTGCTGCCTC	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATAA GACAGCAGT ACAGGAGTCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAACTAATG TTCATCTTCA TTCCTGCCGA GTCGATGGG GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGCCA AAACGGCCGT TCGGCAAA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATGCGCTA AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGGA CATTCCGGCA CAGTCGCAAA GTGCTGCATT GTTTGCCGCA ACAGCGGCAA ACCGGCAAA ACCGGCAAA ACCGGCAAA GTGCTGCATT	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAACCTAATG TTCATCTTCA TTCCTGCGA GCCGAAGGG GCCGAAGGG GCCGAAGGG GCCGAAGGG AAACGGCCGT TCGGCACAA ATGGCTACCAA ATGGGTACGC	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GATGCCGCTA AGCGGTCAG AATGCCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC AAAAATTCAA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTGGCAA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCCATC
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 951 1001	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGGA CATTCCGGCA CAGTCGGCAAA GTGCTGCATT GTTTGCCGCA ACAGCGGCAA ACAGCGCGA ACAGCGGCAA ACAGCGGCAA ACAGCGGCAA GTGCTGCATT GTTTGCCGCA ACAGCGGCGA GATGGAAACG	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA AAGCACCGTC GAGAAATATA GACAGCAGTT ACAGGAGGTCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT GCTTTTAAGGG	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CTTCATCTTCA TTCCTGCGA GCCGAAGGG GCCGAAGGG GCCGAAGGG GCCGAAGGG AAACGCCGT TTGCTGCCA AACGGCCGT TCGGCAGCAA ATGGCTACCC GACTTGGCA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GATGCCGCTA AGCGGTCAG AATGCCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC AAAAATTCAA GAAAAATTCAA	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC ATCCCGTCA ATCCCCGTCA CCTGACGGG ATCTGACGGGG ATCTGACTGACTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCCATC GCGGGGATGT
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1051	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGGA ACCGGCAAA GTGCTGCATT GTTTGCCGCA ACAGCGGCGA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA TTGATTTGCAT GCTTTTAAGGG TTTTAACGGC TTTTACGCC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAACCTAATG TTCATCTTCA TTCCTGCGA GCCGAAGGG GCCGAAGGG GCCGAAGGG AAACGGCCGT TCGGCACAA ATGGCTACT CTGCTGCCA AACGGCCGT CGGCAGCGA ATGGGTACGC GCCTTGGCAA ATGGGTACGC GACTTGGACG CGGCCGGCGA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GATGCCGCTA AGCGGTCAG AATGCCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC AAAAATTCAA GAAAATTCAA GAAAATTCAA GAAAATGCCG	GCGCCACAGG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC ATCCCGTCA ATCCCCGTCA CCTGACGGGG ATCTGACTGGCAAA CTAGAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCATC GCGGGGATGT GGAAAATACA
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 901 1001 1051 1101	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT TTGGATGAAG TGAACGAAT ATTTGGTTGC ATTTATAAAG TGCACGGCAA ACCGGCAAA ACCGGCAAA GTGCTGCATT GTTTGCATT GTTTGCATT CTTTGCTTC ATTCCGGCA AACCGGCAAA ACCGGCAAA ACCGGCAAA GTGCTGCATT GTTTGCCGCA CATGGAAACG GTTCCGGAAGG GCTATCCCCC	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTTGCTA GACCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC TACGCTGATT ATATCTTCGC AAGCGAAAT ATATCTTCGC AGGCGAAAT TTCATACGGA AAAGTCGAT TGATTTGCAT GCTTTAAGGG TTTTAACGCC GACAGATGCC GACAGATGCC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCATCTTCA TTCCTGCCGA GCCGAAGGG GCCGAAGGG GCCGAAGGG GCCGAACGT CTTGCTGCCA AAACGGCCT TCGGCAGCA ATGGGTACGC GACTTGCACCC CGGCCGCCGA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATAGG TTTGAAAATT GAAAAGCGAT GAACACAA TCTGCGCGAT GATGCCGCTA AAGCGGTCAG AATTACCGGT TCGCCTCCGT CGGCCGTGTA CCGTACCGA ATCTGTGGAC AATTACCGG ATCTGTGGAC AAAAATTCAA GAAAATTCAG GAAAATTCGCG GAAAATTCGCG GAAAATTCGCG GAAATTCGGCGT	GCGCCACAGG CCAAGATATG CCAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATT CCGTCGCAAA TGATAATTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCG AGCGCATC GGCGTTCTACGGGGAA CTAGAGGCAG CTGACGGGGA CTAGAGGCAG GGCATTATCG AGCGCCATC GCGGGGATGT GGAAAATACA GTTTGCCGGC
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1101 1151	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGAA CATTCCGGCA CAGGGCGAA ACCGGCAAA GTGCTGCATT GTTTGCCGCA ACAGCGGCA GATGGAAACG GTTCCGGAAGG GTTCCGGAAGG GTTCCGGAAGG GTTCCGGAAGG GCTATCGCCC AAAAAAAGAGC	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCACCGT GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC AAGTCCGC AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT GCTTTAAGGG TTTTACGGCC GACAGATGCC AGGAGATGCC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCATCTTCA TTCCTGCCGA GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGCCA AAACGGCCGT TCGGCAGCAA ATGGGTACGC GACTTGGACG GACTTGGACG GACTTGGACG CGGCCGCGA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGAATGG TTTGAAAATT GAAAAGCGAT GAACCAAC TCTGCGCGAT GATGCCGCTA AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGAC AATTACCGG AATTACCGGA CGGACCACAA GAAAATTCAA GAAAATTCAA GAAAATTCACGGT GGAGCCACAA	GCGCCACAGG CCAAGATATG CCAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCGCCATC GCGGGGATGT GGAAAATACA GTTTGCCGGC ACGACGACG
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 901 1001 1051 1101	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAAAGTT TTGGATGAAG ATTTGGTTGC ATTTATAAAG TGCACGGTCG ATCAGGCGCA ACCGGCAAA GTGCTGCATT GTTTGCCCCA GATGCACC GATGGAAAC GTTTCCGGCA GTTTGCCGCA GTTCCGCAAC GTTTCCGGCAAC GTTTCCGGCAAC GTTTCCGGAAGC TTCCGGAAGC GTTACCCCC AAAAAAGAGC TGTTAAAAAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC AAGTCGCAAATG TTCATACGGA TACATTCGAT GCTTTAAGGG TTTTACGGC GACAGATGCC GACAGATGCC GACAGATGCC GACAGATGCC AGGATGGAT GGTTTAACGCC GACAGATGCC GACAGATGCC AGGATGGAT CGTGCCACTG	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CAAGCTAATG TTCATCTTCA TTCCTGCCGA GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGCCA AAACGGCCGT TCGGCAGCAA ATGGGTACGC GACTTGGACG GACTTGGACG CGGCCGCGA GAAAAGGGCCG CGGCCGCGA GAAAAGGGCCG CGGAGGAGGA TGGCCATTGC TGCCAGTGCACA TTCGCCAGCCACA ATGGCTACCC CGGCCGCCACA CAAAAGGGCCG CGGCCGCGA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCCGC GCGAATGGCG GATTGATAGG TTTGAAAATT GAAAAGCGAT GAACACAA TCTGCGCGAT GATGCCGCTA AAGCGGTCAG AATTACCGGT TCGCCTCCGT CGGCCGTGTA CCGTACCGA ATCTGTGGAC AATTACCGG ATCTGTGGAC AAAAATTCAA GAAAATTCAG GAAAATTCGCG GAAAATTCGCG GAAAATTCGCG GAAATTCGGCGT	GCGCCACAGG CCAAGATATG CCAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCGAA CTAGAGGCGAA CTAGAGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCATC GCGGGGATGT GGAAAATACA GTTTGCCGGC ACGACGACGA ACAATGGCC ACGACGACGA
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 901 901 1001 1105 1101 1151 1201	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGTT ATATAACGTT TTGGATGAAG TGAACGAAT ATTTGGTTGC ATTTATAAAG TGCACGGCGA ACCGGCGAA ACCGGCAAT GTTTGCCGCA ACAGCGGCGA ACAGCGGCA ACAGCGGCA ACAGCGGCA ACAGCGGCA ACAGCGGCA ACAGCGGCCA ACAGCGGCCA ACAGCGGCCA ACAGCGGCCA ACAGCGGCCA ACAGCGGCCA ACAGCAGCCCC AAAAAAGAAC TTTTAAAAA AAGAAATCAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCGTC GAGAAATATA GACAGCAGTT ACAAGTCCGC AGGAGGTCGC TACGCTGATT ATATCTTCGC AAATTGCCG AGGCGAAATG TTCATACGGA TTCATACGGA TTGATTTAAGGC TTTTAAGGC GTTTTAAGGC GACAGATGCG GACAGATGCG AGGATGGATC GCTGCCACTG CGGTTTCAAA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGCCGATT AAAATCAGAA AGAAAGATG CTCATCTCA TTCATCTTCA TTCCTGCCGA GCCGAAGGG GCCGAAGGG GCGGATCGTA CTTGCTGCCA AAACGGCCGT TCGGCAGCA ATGGGTACGC GACTTGGACG CGCCGGCGA GAAAAGGGCC CGGAGGGGAGG	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCGC GCGAATGGCG GATTGATGGG CTTGTAATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATGCCGCTA AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC AATTACAA GAAAATTCAA GAAAATTCAA GAAAATTCAA GAAAATTCAA GAAAATTCAC GGAAGTGGCG GATTCGGCGT GGAGCCACAA TGCTGCCTAC	GCGCCACAGG CCAAGATATG CCAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCGAA CTAGAGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCCATC GCGGGGATGT GGAAAATACA GTTTGCCGGC ACGACGACGA ACAATGGCC CATTGATGAA
40 45 50 55 60	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 901 901 1001 1101 1151 1201 1251	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGCAA ACCGGCAAA GTGCTGCAT GTTTGCCGCA ACAGCGCCA ACAGCGGCAA ACCGGCAAA GTGCTCGCC ACAGCAGCGCAA ACCGGCAAA GTTTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGCAAAAAAAAAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCAGTC GAGAAATATA ACAAGTCCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT GCTTTAAGGG TTTTAACGCC GACAGATGCC GACAGATGCC GGCGAATCC GCTGTTCAAAAAA TTACCAAAAAA GGTCTGGCTC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CCAGCTAATG TTCCTGCCGA GTCGATGGGG GCCGGAAGGG GCGGATCGTA AAACGGCCGT TCGGCAGCA ATGGGTACGC GACTGGACG GCGCGGCGACGA ATGGGTACGC GACTTGGACG GACTTGGACG GACTTGGACG GCGCGGCGACGA ATGGCCGCACG GACAGGCCGA GAAAAGGCC CGCGGAGGAGA TGGCCATTGC GCTGGAGGAGA AGACGCAACT TGAAAAAAGT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCGC GCGAATGGCG GATTGATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCGA ATCTGTGGAC AAAATTCAA GAAAATCGCG GAAGTGCCG GATTCGGCG GATTCGGCG GATTCGGCG TGCGCTACCAA TCTGTGCACCAA CGAACTCACAA TCTGTGCACCAA CGAACTCACCAA CGCACAAA TGCTGCCTACCAA CGCACCAAA TGCTGCCTACCAA CCCATCTACGA GCAGCCCATAC	GCGCCACAGG CCAAGATATG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCCATC GCGGGGATATCG AGCCGCCATC GCAGACAGCGC ACTTGCCGCC ACTTGCCGCC ACTTGCCGCC ACGACACACACACACACACACACACACACACAC
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 901 901 1001 1101 1151 1201 1251 1301	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGCAA ACCGGCAAA GTGCTGCAT GTTTGCCGCA ACAGCGCCA ACAGCGGCAA ACCGGCAAA GTGCTCGCC ACAGCAGCGCAA ACCGGCAAA GTTTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGGAACAC TTCCGCAAAAAAAAAA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCAGTC GAGAAATATA ACAAGTCCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT GCTTTAAGGG TTTTAACGCC GACAGATGCC GACAGATGCC GGCGAATCC GCTGTTCAAAAAA TTACCAAAAAA GGTCTGGCTC	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CCAGCTAATG TTCCTGCCGA GTCGATGGGG GCCGGAAGGG GCGGATCGTA AAACGGCCGT TCGGCAGCA ATGGGTACGC GACTGGACG GCGCGGCGACGA ATGGGTACGC GACTTGGACG GACTTGGACG GACTTGGACG GCGCGGCGACGA ATGGCCGCACG GACAGGCCGA GAAAAGGCC CGCGGAGGAGA TGGCCATTGC GCTGGAGGAGA AGACGCAACT TGAAAAAAGT	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCGC GCGAATGGCG GATTGATAGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT GATTGACGGT AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCCGA ATCTGTGGAC AAAAATTCAA GAAAATTCAA GAAAATGGCG GGATCGGCGT GGAGCCACAA TGCTGCCTAC CGAGCCCACAA TCCTGCCTAC	GCGCCACAGG CCAAGATATG CCAAGATATG CAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG GGCATTATCG AGCCGCCATC GCGGGGATATCG AGCCGCCATC GCAGACAGCGC ACTTGCCGCC ACTTGCCGCC ACTTGCCGCC ACGACACACACACACACACACACACACACACAC
40 45 50 55 60	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1101 1101 1151 1201 1251 1301	ATGGCTAGCC TCCTGTTGTT CAGGTTCTCA GCGGCAGTTT CAAACCCAAA CCGCCGAATC GATTCCGCCC TGGAAGGGTT ATATAACGTT TTGGATGAAG TGAACGAATT ATTTGGTTGC ATTTATAAAG TGCACGGCAA ACCGGCAAA GTGCTGCAAA GTGCTGCAAT GTTTGCCGCA ACAGCGCCAA ACCGGCAAA GTGCTCGCC ACAGCGGCAAA GTTTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACG TTCCGGAACA CGACTTTAAAA AAGAAATCAA GACGGCACAA CGACTTTAAA CCGTCAATGA	GCTGAAAAAG AGGACAGGGC CGGCAGAAAA AATGAAGACG CGCAAATCAA CCGCGTCAAA GATTTGGCTA GACCCACTGT AAGCACCAGTC GAGAAATATA ACAAGTCCGC TACGCTGATT ATATCTTCGC AAATTGCCCG AGGCGAAATG TTCATACGGA AAAGTCGATT TGATTTGCAT GCTTTTAAGGG TTTTAACGCC GACAGATGCC AGGATGCCACTC CGCTTTCAAA TTACCAAAAA TTACCAAAAA CAAAACAA	AGACAGAGGT GCGCCATCCA TACAGGCAAT AGGGACCGCA ACAGGGAACA CCCTGCACCT ATGGCGTTTT AAAGGCGATT AAAATCAGAA AGAAAGATGG CCAGCTTTC TTCCTGCCGA GTCGATGGG GCCGAAGGG GCGGATCGTA TCGGCAGCAA ATGGGTACT TCGGCAGCAA ATGGGTACGC GACTTGGACG GACTTGGACG GACTTGGACG GACTTGGACG GACTTGGACG GACTTGGACG GACTTGGACG GACTGGACGA ATGGCTGACG GACTGGACGA TGGCCGCGA GAAAAGGCC CGCGCGAGAGA TGGCCATTGC CGCTGGAGAGA AGACGCAACT TGAAAAAAGT AACGTCGATG AACCGAGTTG AACCGAGTTA	AAAAGAAGAT CACAAGGCAG GGCGGTGCGG AAATGATATG ACCAACCGC GCGAATGGCG GATTGATGG TTTGAAAATT GAAAAGCGAT GAACTAACAA TCTGCGCGAT AAGCGGTCAG AATTACCGGT TGCCCTCCGT CGGCCGTGTA CCGTACCGA ATCTGTGGAC AAAATTCAA GAAAATCGCG GAAGTGCCG GATTCGGCG GATTCGGCG GATTCGGCG TGCGCTACCAA TCTGTGCACCAA CGAACTCACAA TCTGTGCACCAA CGAACTCACCAA CGCACAAA TGCTGCCTACCAA CGCACCAAA TGCTGCCTACCAA CCCATCTACGA GCAGCCCATAC	GCGCCACAGG CCAAGATATG CCAACAACGGA CCGCAAAATT CGATTCTTCA GTAGCAATTT CCGTCGCAAA TGATAATTTA TAAATGAGTC AAATTTACTA ATATGTCATC TCAGGCGTTC ATCCCCGTCA CCTGACGGGG ATCTGACTTA GTGCAAGGCG CAACGGCGAA CTAGAGGCAG CGGGGATTATCG AGCCGCCATC GCGGGGATGT GGAAAATACA GTTTGCCGGC ACGCGCAC CTGACGGC ACTTACC CCGGGGATGT CGAAAATACA GTTTGCCGCC ACGACGACGA CTGACCAAAA ACCTGCAGAA ATGCCCATT

	1 601	maaaaaaaaa	m2 m2 2 ac2 a2	mmmccmcz z c	202002200	3 3 3 M 3 M C C M 3
	1601			TTTGCTGAAG		
	1651			AGCCGTGGCT		
	1701			CCGATTCATT		
5	1751	CAGACGAAGC			CCAAACAGAC	
3	1801	ACCAAACAAA			GCTGCAGAAA	
	1851	CAAAGCCGAA			TACTGCAGCC	
	1901	AAGCTGTCGC		ACCGACATCA		
	1951			AGCAAACAGT		
1.0	2001			TCAGAATTGA		
10	2051	AAAAATTGGA		GCTTCTGCTG	AAAAATCCAT	TGCCGATCAC
	2101	GATACTCGCC	TGAACGGTTT	GGATAAAACA	GTGTCAGACC	TGCGCAAAGA
	2151	AACCCGCCAA	GGCCTTGCAG	AACAAGCCGC	GCTCTCCGGT	CTGTTCCAAC
	2201	CTTACAACGT	GGGTCGGTTC	AATGTAACGG	CTGCAGTCGG	CGGCTACAAA
	2251	TCCGAATCGG	CAGTCGCCAT	CGGTACCGGC	TTCCGCTTTA	CCGAAAACTT
15	2301	TGCCGCCAAA	GCAGGCGTGG	CAGTCGGCAC	TTCGTCCGGT	TCTTCCGCAG
	2351	CCTACCATGT	CGGCGTCAAT	TACGAGTGGT	AACTCGAG	
	1	MASPDVKSAD	TLSKPAAPVV	AEKETEVKED	APQAGSQGQG	APSTQGSQDM
	51	AAVSAENTGN	GGAATTDKPK	NEDEGPQNDM	PQNSAESANQ	TGNNQPADSS
20	101	DSAPASNPAP	ANGGSNFGRV	DLANGVLIDG	PSQNITLTHC	KGDSCNGDNL
	151	LDEEAPSKSE	FENLNESERI	EKYKKDGKSD	KFTNLVATAV	QANGTNKYVI
	151 201			EKYKKDGKSD RRSLPAEMPL		-
		IYKDKSASSS	SARFRRSARS		IPVNQADTLI	VDGEAVSLTG
	201	IYKDKSASSS HSGNIFAPEG	SARFRRSARS NYRYLTYGAE	${\tt RRSLPAEMPL}$	IPVNQADTLI VQGEPAKGEM	VDGEAVSLTG LAGTAVYNGE
25	201 251	IYKDKSASSS HSGNIFAPEG	SARFRRSARS NYRYLTYGAE PYPTRGRFAA	RRSLPAEMPL KLPGGSYALR	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI
25	201 251 301	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG
25	201 251 301 351	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE
25	201 251 301 351 401	TYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE
25	201 251 301 351 401 451	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT SEIEKLTTKL	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK ADTDAALADT	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY GLGLKKVVTN	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ LNKLGENITT	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE FAEETKTNIV
25 30	201 251 301 351 401 451 501	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT SEIEKLTTKL KIDEKLEAVA	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK ADTDAALADT DTVDKHAEAF	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY GLGLKKVVTN DAALDATTNA	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ LNKLGENITT NTKADEAVKT	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE FAEETKTNIV ANEAKQTAEE
	201 251 301 351 401 451 501 551	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT SEIEKLTTKL KIDEKLEAVA TKQNVDAKVK	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK ADTDAALADT DTVDKHAEAF AAETAAGKAE	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY GLGLKKVVTN DAALDATTNA NDIADSLDET	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ LNKLGENITT NTKADEAVKT DKAEAVAAKV	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE FAEETKTNIV ANEAKQTAEE TDIKADIATN
	201 251 301 351 401 451 501 551 601	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT SEIEKLTTKL KIDEKLEAVA TKQNVDAKVK KDNIAKKANS	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK ADTDAALADT DTVDKHAEAF AAETAAGKAE ADVYTREESD	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY GLGLKKVVTN DAALDATTNA NDIADSLDET AAAGTANTAA SKFVRIDGLN	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ LNKLGENITT NTKADEAVKT DKAEAVAAKV ATTEKLDTRL	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE FAEETKTNIV ANEAKQTAEE TDIKADIATN ASAEKSIADH
	201 251 301 351 401 451 501 551 601 651	IYKDKSASSS HSGNIFAPEG VLHFHTENGR DGNGFKGTWT KKEQDGSGGG DGTITKKDAT SEIEKLTTKL KIDEKLEAVA TKQNVDAKVK KDNIAKKANS DTRLNGLDKT	SARFRRSARS NYRYLTYGAE PYPTRGRFAA ENGGGDVSGR GATNDDDVKK AADVEADDFK ADTDAALADT DTVDKHAEAF AAETAAGKAE ADVYTREESD VSDLRKETRQ	RRSLPAEMPL KLPGGSYALR KVDFGSKSVD FYGPAGEEVA AATVAIAAAY GLGLKKVVTN DAALDATTNA NDIADSLDET AAAGTANTAA	IPVNQADTLI VQGEPAKGEM GIIDSGDDLH GKYSYRPTDA NNGQEINGFK LTKTVNENKQ LNKLGENITT NTKADEAVKT DKAEAVAAKV ATTEKLDTRL LFQPYNVGRF	VDGEAVSLTG LAGTAVYNGE MGTQKFKAAI EKGGFGVFAG AGETIYDIDE NVDAKVKAAE FAEETKTNIV ANEAKQTAEE TDIKADIATN ASAEKSIADH NVTAAVGGYK

	ELISA	Bactericidal
ΔG287-953-His	3834	65536
ΔG287-961-His	108627	65536

The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens (using 287-GST) for 919 and ORF46.1:

	Mixture with 287	Hybrid with ΔG287
919	32000	128000
ORF46.1	128	16000

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained:

WO 01/64922

	9	19	ORF46.1		
Strain	Mixture	Hybrid	Mixture	Hybrid	
NGH38	1024	32000	-	16384	
MC58	512	8192	-	512	
BZ232	512	512	-	_	
MenA (F6124)	512	32000	-	8192	
MenC (C11)	>2048	>2048	-	-	
MenC (BZ133)	>4096	64000	-	8192	

The hybrid proteins with $\Delta G287$ at the N-terminus are therefore immunologically superior to simple mixtures, with $\Delta G287$ -ORF46.1 being particularly effective, even against heterologous strains. $\Delta G287$ -ORF46.1K may be expressed in pET-24b.

The same hybrid proteins were made using New Zealand strain 394/98 rather than 2996:

5	∆G287NZ	-919				
	1	ATGGCTAGCC	CCGATGTCAA	GTCGGCGGAC	${\tt ACGCTGTCAA}$	AACCTGCCGC
	51	CCCTGTTGTT	TCTGAAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCCACAGG
	101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCG	CACAAGGCGG	TCAAGATATG
	151	GCGGCGGTTT	CGGAAGAAAA	TACAGGCAAT	GGCGGTGCGG	CAGCAACGGA
10	201	CAAACCCAAA	AATGAAGACG	AGGGGGCGCA	AAATGATATG	CCGCAAAATG
	251	CCGCCGATAC	AGATAGTTTG	ACACCGAATC	ACACCCCGGC	${\tt TTCGAATATG}$
	301	CCGGCCGGAA	ATATGGAAAA	CCAAGCACCG	GATGCCGGGG	AATCGGAGCA
	351	GCCGGCAAAC	CAACCGGATA	TGGCAAATAC	GGCGGACGGA	ATGCAGGGTG
	401	ACGATCCGTC	GGCAGGCGGG	GAAAATGCCG	GCAATACGGC	TGCCCAAGGT
15	451	ACAAATCAAG	CCGAAAACAA	TCAAACCGCC	GGTTCTCAAA	ATCCTGCCTC
	501	TTCAACCAAT	CCTAGCGCCA	CGAATAGCGG	TGGTGATTTT	GGAAGGACGA
	551	ACGTGGGCAA	TTCTGTTGTG	ATTGACGGGC	CGTCGCAAAA	${\tt TATAACGTTG}$
	601	ACCCACTGTA	AAGGCGATTC	TTGTAGTGGC	AATAATTTCT	TGGATGAAGA
	651	AGTACAGCTA	AAATCAGAAT	TTGAAAAATT	AAGTGATGCA	GACAAAATAA
20	701	GTAATTACAA	GAAAGATGGG	AAGAATGACG	GGAAGAATGA	TAAATTTGTC
	751	GGTTTGGTTG	CCGATAGTGT	GCAGATGAAG	GGAATCAATC	${\tt AATATATTAT}$
	801	CTTTTATAAA	CCTAAACCCA	CTTCATTTGC	GCGATTTAGG	CGTTCTGCAC
	851	GGTCGAGGCG	GTCGCTTCCG	GCCGAGATGC	CGCTGATTCC	CGTCAATCAG
	901	GCGGATACGC	TGATTGTCGA	TGGGGAAGCG	GTCAGCCTGA	CGGGGCATTC
25	951	CGGCAATATC	TTCGCGCCCG	AAGGGAATTA	CCGGTATCTG	ACTTACGGGG
	1001	CGGAAAAATT	GCCCGGCGGA	TCGTATGCCC	TCCGTGTTCA	AGGCGAACCT
	1051	TCAAAAGGCG	AAATGCTCGC	GGGCACGGCA	GTGTACAACG	GCGAAGTGCT
	1101	GCATTTTCAT	ACGGAAAACG	GCCGTCCGTC	CCCGTCCAGA	GGCAGGTTTG
	1151	CCGCAAAAGT	CGATTTCGGC	AGCAAATCTG	TGGACGGCAT	TATCGACAGC
30	1201	GGCGATGGTT	TGCATATGGG	TACGCAAAAA	TTCAAAGCCG	CCATCGATGG
	1251	AAACGGCTTT	AAGGGGACTT	GGACGGAAAA	TGGCGGCGGG	GATGTTTCCG
	1301	GAAAGTTTTA	CGGCCCGGCC	GGCGAGGAAG	TGGCGGGAAA	ATACAGCTAT
	1351	CGCCCAACAG	ATGCGGAAAA	GGGCGGATTC	GGCGTGTTTG	CCGGCAAAAA
~ ~	1401	AGAGCAGGAT	GGATCCGGAG	GAGGAGGATG	CCAAAGCAAG	AGCATCCAAA
35	1451	CCTTTCCGCA	ACCCGACACA	TCCGTCATCA	ACGGCCCGGA	CCGGCCGGTC
	1501	GGCATCCCCG	ACCCCGCCGG	AACGACGGTC	GGCGGCGGCG	GGGCCGTCTA
	1551	TACCGTTGTA	CCGCACCTGT	CCCTGCCCCA	CTGGGCGGCG	CAGGATTTCG
	1601	CCAAAAGCCT	GCAATCCTTC	CGCCTCGGCT	GCGCCAATTT	GAAAAACCGC
	1651	CAAGGCTGGC	AGGATGTGTG	CGCCCAAGCC	TTTCAAACCC	CCGTCCATTC
40	1701	CTTTCAGGCA	AAACAGTTTT	TTGAACGCTA	TTTCACGCCG	TGGCAGGTTG
	1751	CAGGCAACGG	AAGCCTTGCC	GGTACGGTTA	CCGGCTATTA	CGAGCCGGTG
	1801	CTGAAGGGCG	ACGACAGGCG	GACGGCACAA	GCCCGCTTCC	CGATTTACGG
	1851		GATTTTATCT		GCCTGCCGGT	TTGCGGAGCG
	1901	GAAAAGCCCT	TGTCCGCATC	AGGCAGACGG	GAAAAAACAG	CGGCACAATC
45	1951	GACAATACCG	GCGGCACACA	TACCGCCGAC	CTCTCCCGAT	TCCCCATCAC
	2001	CGCGCGCACA	ACGGCAATCA	AAGGCAGGTT	TGAAGGAAGC	CGCTTCCTCC

WO 01/64922

	2051	CCTACCACAC	GCGCAACCAA	ATCAACGGCG	GCGCGCTTGA	CGGCAAAGCC
	2101		GTTACGCCGA			
	2151		GGCCGTCTGA			
	2201	• •	CAAAAACGAA		- '	
5	2251	•	GCTACCTCAA			
J	2301		CGGCAAAATC			
	2351		TATCTTTTTC			
	2401		CACTGGGCAC			
			TACATTACCT			
10	2451	_				
10	2501		CCGCAAAGCC			
	2551	•	TTAAAGGCGC			
	2601	•	GGCGAACTTG			
	2651	GGCAGCTCCT	ACCCAACGGT	ATGAAGCCCG	AATACCGCCC	GTAAAAGCTT
1.5						
15	1		TLSKPAAPVV			
	51		GGAAATDKPK			
	101	PAGNMENQAP	DAGESEQPAN	QPDMANTADG	MQGDDPSAGG	ENAGNTAAQG
	151	TNQAENNQTA	GSQNPASSTN	PSATNSGGDF	GRTNVGNSVV	IDGPSQNITL
	201	THCKGDSCSG	NNFLDEEVQL	KSEFEKLSDA	DKISNYKKDG	KNDGKNDKFV
20	251	GLVADSVQMK	GINQYIIFYK	PKPTSFARFR	RSARSRRSLP	AEMPLIPVNQ
	301	ADTLIVDGEA	VSLTGHSGNI	FAPEGNYRYL	TYGAEKLPGG	SYALRVQGEP
	351	SKGEMLAGTA	VYNGEVLHFH	TENGRPSPSR	GRFAAKVDFG	SKSVDGIIDS
	401		FKAAIDGNGF			
	451		GVFAGKKEOD			
25	501		GGGGAVYTVV			
20	551		FQTPVHSFQA			
	601		ARFPIYGIPD			
			LSRFPITART			
	651					
30	701		ELFFMHIQGS			
30	751		TSMQGIKAYM			
	801		GEYAGAVDRH			_
	851	GSAIKGAVRV	DYFWGYGDEA	GELAGKQK'I"I'	GYVWQLLPNG	MKPEYRP*
25						
35	<u>ΔG287NZ</u>					
35	1	ATGGCTAGCC	CCGATGTCAA			
35	1 51	ATGGCTAGCC CCCTGTTGTT	TCTGAAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCCACAGG
35	1	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA	TCTGAAAAAG AGGACAGGGC	AGACAGAGGC GCGCCATCCG	AAAGGAAGAT CACAAGGCGG	GCGCCACAGG TCAAGATATG
	1 51	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA	AGACAGAGGC GCGCCATCCG TACAGGCAAT	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG	GCGCCACAGG TCAAGATATG CAGCAACGGA
35 40	1 51 101	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT	TCTGAAAAAG AGGACAGGGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG	GCGCCACAGG TCAAGATATG CAGCAACGGA
	1 51 101 151	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG
	1 51 101 151 201	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG
	1 51 101 151 201 251	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGGAA GCCGGCAAAC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG
	1 51 101 151 201 251 301	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGGAA GCCGGCAAAC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG
	1 51 101 151 201 251 301 351	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGGAA GCCGGCAAAC ACGATCCGTC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT
40	1 51 101 151 201 251 301 351 401 451	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGGAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC
40	1 51 101 151 201 251 301 351 401 451 501	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGGAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA
40	1 51 101 151 201 251 301 351 401 451 501 551	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGCCGAAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATACCGC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG
40	1 51 101 151 201 251 301 351 401 451 501 551 601	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCCACTGTA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAGGCGATTC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG ATTGACGGGC TTGTAGTGGC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 651	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA ACTACAGCTA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAGGCGATTC AAATCAGAAT	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG ATTGACGGGC TTGTAGTGGC TTGAAAATT	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGA GCGAACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA
40	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGATAC GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA AGTACAGCTA GTAATTACAA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAACAA CCTAGCGCCA TTCTGTTGTG AAGGCGATTC AAATCAGAAT GAAAGATGGG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG CCGAATAGCGC CGAATAGCGG ATTGACGGC TTGTAGTGGC TTGAAAATT AAGAATGACG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGCA	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGAAA CCGGCCGAAA ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GTAATTACAA GGTTTGGTTG	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACT ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAACAA TTCTGTTGTG AAGGCGATTC AAATCAGAAT GAAAGATGGC CCGATAGTGCC CAAACAA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TCGAAATGCCG TCAAACCGCC CGAATACCGG ATTGACGGC TTGTAGTGGC TTGAAAAATT AAGAATGACG GCAGATGACG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GCCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAAGAATGA GGAATCAATC	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGGAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GTAATTACAA GGTTTGGTTG CTTTTATAAA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACT ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA TTCTGTTGTG AAGCGCATTCGTTGTG AAGCGATTCCA AATCAGAAT GAAAGATGGC CCGATAGCGC CCGATAGCGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGGC ATTGACGGCC TTGTAGTGGC TTGAAAAATT AAGAATGACG GCAGATGAAG CTTCATTGC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAAGAATGA GGAATCAATC GCGATTTAGG	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC
40 45	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 851	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGGAA GCCGGCAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GGTACTTGGTTC GTAATTACAA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA TTCTGTTGTG AAGCGCCA TTCTGTTGTG AAGCGGATA GAAACAATCAGAAT GCAAGATGGC CCGATAGTGT CCTAAACCCA GTCGCTTCCG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TCGCAAATGCCG TCAAACCGCC CGAATAGCGGC TTGACAGGGC TTGACAGGC TTGACAGGC TTGAAAATT AAGAATGACG GCAGATGACG GCAGATGACG GCAGATGACG GCAGATGACG GCCGAGATGACG GCCGAGATGACG GCCGAGATGC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAAGAATGA GGAATCAATC GCGATTTAGG CGCTGATTCC	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGGAAAC CCGGCCGAAA ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GGTATTGCTTG CTTTTATAAA GGTCGAGGCG GCGGATACGC GCGGATACGC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAGTCAGATC AAATCAGAAT GAAAGATGGC CCGATAGTGT CCTAAACCCA GTCGCTTCCG TGTTCCG	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TCGAAATAC GCAATACCGC CGAATACCGC TTGACGGC TTGACGGC TTGAAAATT AAGAATGACG GCAGATGAAG CTTCATTTGC GCCGAGATGC TGGGGAAGCC TGGGGAAGCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAAGAATGA GGAATCAATC GCGATTTAGG CGCTGATTCC GTCAGCCTGA	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGGGGCATTC
40 45	1 51 101 151 201 251 301 351 401 451 501 651 701 751 801 851 901	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAAC ACGATCCGTC ACAAATCAAG TCAACCAAT ACGTGGCCAA ACCCACTGTA AGTACAGCTA GGTACAGCTA GGTATACAA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGGATACGC CGGCAATATC	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CTTAGCGCCA TTCTGTTGTG AAGGCGATTC AAATCAGAAT GAAAGATGGG CCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGCGA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG ATTGACGGGC TTGAAAATT AAGAATGACG GCAGATGAAG CTTCATTTGC GCCGAGATGC TGGGGAAGCG AGGGGAAGCG AAGGGAATTA	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAATCAATC GCGATTTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCTG	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGGATACCC CGGCAATATC CGGAAAAATT CGGAAAAATT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAGGCGATTC AAATCAGAAT GAAAGATGGG CCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGCGA TTCGCGCCCG GCCCGGCGGA	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG ATTGACGGGC TTGAAAATT AAGAATGACG GCAGATGACG CTTCATTTGC GCCGAGATGC TGGGGAAGCG AGGGAATTA TCGTATGCCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAATCAAT GCGATTTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCTG TCCGTGTTCA	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA TAGAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 951 1001	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGGATACC CGGCAATATC CGGAAAAATT TCAAAAGGCG	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAGGCGATTC AAATCAGAAT GAAAGATGGG CCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGCGA TTCGCGCCCG GCCCGGCGGA AAATGCTCGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GAAAATGCCG CGAATAGCGGC ATTGACGGGC TTGAAAATT AAGAATGACG GCAGATGAAG CTTCATTTGC GCCGAGATGC TGGGGAAGCG AGGGAATTA TCGTATGCCC GGGCACGCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAATCAAT GCGATTTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCTG TCCGTGTTCA GTGTACAACG	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA TAGATGAAGA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 851 901 951 1001 1051	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCAAC ACGATCAGT ACGATCAGT ACGATCAAT ACGTGGCAA ACCACTGTA AGTACAGCTA GGTATTTCAA GGTTTGGTTG CTTTTATAAA GGTCGAGCG CCGCAATATC CCGCAATATC CGGAAAAATT CAAAAGGCG GCATTTCAT TCAAAAGGCG GCATTTTCAT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CCAACCGGATA CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAAGCATGTTC AAATCAGAAT CCTAAGCGCATC CCTAAACCCA GTCGCTTCCG TGATTGTCGA TTCGCGCCCC GCCCGCCGCAAAACCA AAATGCTCGC ACGGAAAACCC ACGGAAAACCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATAGCGG ATTGACGGG TTGAAAATT AAGAATGACG CTTCATTTGC GCCGAGATGC TGGGGAAGCG AGGGAATTA TCGTATGCCC GGGCACGGCA GCCGTCCGTC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAAGAATGA GGAATCAAC CGCTGATTCC GTCAGCCTGA CCGGTATCTG TCCGTGTTCA GTGTACAACG CCCGTCCAGA	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAATAA TAAATTTGTC AATATATAT CGTTCTGCAC CGGCGAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT GCGAAGTGCT GCGAAGTGCT
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 651 701 751 801 851 901 951 1001 1101 1151	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAA GCCGGCAAAC ACGATCAGC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCACTGTA AGTACAGCTA GTAATTACAA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGAATATC CCGGAAAATT TCAAAAGGCG GCATTTTCAT CCGCAAAAGT CCGCAAAAGT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA CCTAGCGCCA TTCTGTTGTG AAGCGATTC AAATCAGAAT GAAAGATGGCCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGTCGC TGCTTCCG TGATTGTCGCA AATCAGAACA GCCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGTCGA AATCCGCCCG GCCCGGCGGA AAATGCTCGC ACGGAAAACG CGATTTCGGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAATGCCG CGAATAGCGGC ATTGACGGGC ATTGACGGGC TTGTAGTGGC TTGAAAATT AAGAATGACG GCAGATGACG GCCGAGATGC GCCGAGATGC TGGGAAGCC GCGGAATGC GCCGAGATGC TGGGGAAGCC AAGGGAATTC TGGGGAAGCC AGGGAACCC GGGCACGGCA GCCGTCCGTC AGCAAATCTG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATAAGC GGATTAGG CGCTGATTCC GTCGCATTCC GTCGCATTCC GTCGCATTCC GTCGCTGATTCC GTCGCTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTCAGA TGGACGCAT TGGACGCAT	GCGCCACAGG TCAAGATATG CAGCAACAGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTGACGGG AGGCGAACCT GCGAAGTGCT GCGAAGTGCT GCGAAGTGCT TATCGACACC
40 45 50	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 851 901 951 1001 1051	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGAAA GCCGGCAAAC ACGATCAGC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCACTGTA AGTACAGCTA GTAATTACAA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGAATATC CCGGAAAATT TCAAAAGGCG GCATTTTCAT CCGCAAAAGT CCGCAAAAGT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CCAACCGGATA CCGAAAACAA CCTAGCGCCA TTCTGTTGTG AAAGCATGTTC AAATCAGAAT CCTAAGCGCATC CCTAAACCCA GTCGCTTCCG TGATTGTCGA TTCGCGCCCC GCCCGCCGCAAAACCA AAATGCTCGC ACGGAAAACCC ACGGAAAACCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAATGCCG CGAATAGCGGC ATTGACGGGC ATTGACGGGC TTGTAGTGGC TTGAAAATT AAGAATGACG GCAGATGACG GCCGAGATGC GCCGAGATGC TGGGAAGCC GCGGAATGC TGGGGAAGCC TGGGGAAGCC TGGGGAACCG AAGGGAATTC TGGGCACGCCA GCCGTCCGTC AGCAAATCTG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATAAGC GGATTAGG CGCTGATTCC GTCGCATTCC GTCGCATTCC GTCGCATTCC GTCGCTGATTCC GTCGCTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTTCA TCCGTGTCAGA TGGACGCAT TGGACGCAT	GCGCCACAGG TCAAGATATG CAGCAACAGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTGACGGG AGGCGAACCT GCGAAGTGCT GCGAAGTGCT GCGAAGTGCT TATCGACACC
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 651 701 751 801 851 901 951 1001 1101 1151	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGCCGATAC CCGGCCGAAA GCCGCCAAA ACCAATCAAG TTCAACCAAT ACGTGGCAA ACCACTGTA AGTACAGCTA GGTATTGGTTG CTTTTATAAA GGTCGAGGCG GCGAATATC CCGCAAAATCAC GCGCAATATC CCGCAAAAGT TCAAAGGCG GCATTTTCAT CCGCAAAAGT CCGCAAAAGT CCGCAAAAGT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA CCTAGCGCCA TTCTGTTGTG AAGCGATTC AAATCAGAAT GAAAGATGGCCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGTCGC TGCTTCCG TGATTGTCGCA AATCAGAACA GCCGATAGTGT CCTAAACCCA GTCGCTTCCG TGATTGTCGA AATCCGCCCG GCCCGGCGGA AAATGCTCGC ACGGAAAACG CGATTTCGGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG CGAATAGCGG ATTGACGGC TTGTAGTGGC TTGAAAATT AAGAATGACG GCAGATGACG GCAGATGACG GCGGAAGTGACGC GGGCACGCC AGGCACGCC AGCGCACGCC GCGCACATC TCGTATTCC CGGCCACGCC AGCACATC TCGTATTCC TCGCCACACACAC GCCGTCCGTC AGCAAATCTG	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATCAATC GCGATTTAGG CGCTGATTCC GTCAGCTGA TCCGGTATCTG TCCGTGTTCA GTGTCAAACG CCCGTCCAGA TGGACGCAT TTCAAAGCCG	GCGCCACAGG TCAAGATATG CAGCAACAGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT GCCAGGTTTG TATCGACAGC CCGTCAATCGC
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 651 701 751 801 851 901 951 1001 1101 1151	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGCCGATAC CCGCCGATAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA AGTACAGCTA GGTATTGGTTG CTTTTATAAA GGTCGAGGCG GCGATACC CGGAAAATT CAAAAGGCG GCATTTCAT CCGCAAAAGT CCGCAAAAGT CCGCAAAAGT CCGCAAAAGT AAACGCCTTT AAACGCCTTT AAACGCCTTT AAACGCCTTT AAACGCCTTT AAACGCCTTT AAACGCCTTT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CCAACCGGATA CCGAAACAA CCTAGCGCCA TTCTGTTGTG AAGCGATTC AAATCAGAAT CCTAAACCA GTCGCTTCCG TGATTGTCGA TTCCGCCCG GCCCGGCGAAAACA AATGCCGCCG ACGGAAAACG CCGATAGTCGC CCGCGCGCGA AAATGCTCGC CCGCTCGCCGCCCCC CCGCTCGCCCCC CCCGCCGCAAAACCC CCGATTTCCGC CCGATTTCCGC CCGATTTCCGC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCCCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG CTGAACCGCC CGAATAGCGG ATTGACGGC TTGTAGTGGC TTGAAAAATT AAGAATGACG GCAGATGACG GCCGAGATGC GCCGAGATGC GGGCACGCC AGCGCACGCCA GCCGTCCGTC AGCAAAAAA GGACGGAAAAA GGACGGAAAAA	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GCCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATCAATC GCGATTTAGG CGCTGATTCC GTCAGCTGA CCGGTATCTC GTCGGTATCTG TCCGTGTTCA GTGTACACG CCCGTCCAGA TGGACGGCAT TTCAAAGCCG TGGCGGCGGG	GCGCCACAGG TCAAGATATG CAGCAACAGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGGCGAACTT ACGTCAACAG AGCGAACCT GCGAAGTGTT GGCAAGTTTG GCAAGTTTG TATCGACAGC CGCAACTTC GCAAGTTTC GCAAGTTTC GCAAGTTTC TATCGACAGC CCATCGATGG GATGTTTCCG
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 651 701 751 801 851 901 901 1001 1101 1151 1201 1251	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGCCGATAC CCGCCGATAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGGCAA ACCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGATACC CGGAAAATT CCGGAAAATT TCAAAAGCCG GCATTTCAT CCGCAAAAGT CCGCAAAAGT CCGCAAAAGT AAACGCCTT AAACGCCTT AAACGCCTT AAACGCCTT AAACGCCTTT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CCAACCGGATA CCGAACAA CCTAGCGCCA TTCTGTTGTG AAGCGATTC AAATCAGAAT CCTAAACCA GCCATAGTGT CCTAAACCA GTCGCTTCCG TGCTTCCG TGCTTCCG TCGCTCCG CCCGCCGCA AAATGCTCGC ACGGAAAACA ACGGAAAACG CCGATTTCGCG CCGCCGCGCGA AAATGCTCGC CGATTTCGCG CCGATTTCGCG ACGGAAAACG CGATTTCGGC CGATTTCGCG CGATTTCGCG ACGGAAAACG CGATTTCGGC CGATTTCGGC CGATTTCGGC ACGGAAAACG AAGGGGACTT	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG ATTGACGGC ATTGACGGC TTGTAGTGGC TTGAAACTT AAGAATGACG GCAGATGACG GCAGATGACG GCGGAATTT TGCGCAAACT TCGTATTCC GCGGAATTAC GGCGCACGCA ACGCAAATTA TCGTATGCC GGGCACGGCA GCCGTCCGTC AGCAAATCTG TACGCAAAAA GGACGGAAAA	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGATGC GGAATCAATC GCGATTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCTG TCCGTGTTCA TCCGTGTTCA TGGACGGCAT TTCAAAGCCG TGGCGGCGGG TGGCGGCGGG	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT GCGAAGTGCT TATCGACAGC CGGAAGTGCT TATCGACAGC CCATCGATGG GATGTTTCCG ATACAGCTAT
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 751 801 851 901 951 1001 1105 1101 1151 1201 1251 1301	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGCCGATAC CCGGCCGATAC CCGGCCGATAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACCTGGGCAA ACCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGATATC CGGAAAATT TCAAAAGCCG GCATTTCAT CCGCAAAAGT CCGCAAAAGT AAACGCCTT AAACGCCTT AAACGCCTT CAAACGCTTT CAAACGCCTT CCGCAAAAGT CCGCAACAG	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAACAA CTTCTGTTGTG AAGCGATTC AAATCAGAAT GAAAGATGGC CCGATAGTGT CCTAAACCA GTCGCTTCCG TGTTGTCGA TTCTGCGCCCG GCCCGGCGA AAATGCTCGC AAATGCTCGC CCGGATATGTCGA TTCGCGCCCG CCCGGCGGA AAATGCTCGC CGATTTCGGC CGGCCCGCCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATC CCAAGCACCG TGGCAAATAC GAAAATGCCG TCAAACCGCC CGAATACCGG ATTGACGGC TTGTAGTGGC TTGAAAAATT AAGAATGACG GCAGATGACG GCGGAATGCC GCGGAATTTTCC GCGGAAATTT TCGTATTCC GCGGAATTA TCGTATTCC GGGCACGCA AGGGAATTA TCGTATCCC GGGCACGCA GCCGTCCGTC AGCAAATCTG TACGCAAAAA GGACGGAAAA GGCGGAAAAA GGCGGAAAAA GGCGGAAAAA	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTTT CGTCGCAAAA AATAATTTCT AAGTGATGATGCA GGAATCAATC GCGATTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCTG TCCGTGTTCA GTGACACG TCGCGCAGA TGGACGGCAT TTCAAAGCCG TGGCGCGGGAAA GGCGTGTTTG	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGTG TGCCCAAGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA GACAAAATAA TAAATTTGTC AATATATTAT CGTTCTGCAC CGGCGAACTCA ACGGGGCATTC ACGGGGCATTC ACGTCAGGGG TATCAGACT GCGAAGTGCT GCGAAGTGCT GCGAAGTGCT TATCGACAGC CCATCGATGG GATGTTTCCG ATACAGCTAT CCGGCAAAAA
40 45 50 55	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 951 1001 1051 1151 1201 1251 1301 1351	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGATAC CCGGCCGAAA ACCAATCAGT ACAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGAATATC CGGAAAATT TCAAAGGCG GCGATATT CAGAAGTTTCAT TCAAAGGCG GCAATATT CAGAAGTT TCAAAGGCAGAT TCAAAGGCAGAT TCAAAGGCAGAT TCAAAAGTTTTCAT AAACGCCTTT AAACGCCTTT CAAAGTTTTA CGCCCAACAG AGAGCAGGAT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA TTCTGTTGTG AAGGCGATT CAAACCGATT CAAACCCA TTCTGTTGTG AAGGCGATT CAAACCCA GTCGCTTCCG TGATTGTCGA TTCGCGCCCG GCCGGCGAAAACA CGATTTCGCG CCCGGCGGAAAACG AAATGCTCGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC AAGGGGACTT CGGCCCGGCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GCAAACCGCC CGAATACCGG ATTGACGGC TTGTAGTGGC TTGAAAATT AAGAATGACG GCAGATGACG GCAGATGACG GCGGAGATGC GCGGAGATGC TGGGGAATTT AGGCAAATT AGGCAAATT CGTATGCC GGGCACGCA GCCGTCCGTC AGCAAAATCT TACGCAAAAA GGCCGAAAAA GGACGGAAAA GGACGGAAAA GGACGGAAAA GGCGAGGAAAC GGCGGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAGAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC GGCGGAAAC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATCAATC GCGATTAGG CGCTGATTCC GTCAGCCTGA TCCGTTCACACG TCCGTCCAGA TGGACGCAT TTCAAAGCCG TGGCGCGGG TGGCGCGGGAAA GGCGTGTTTC CGCGTCTTTC	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGGT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA TAAATTTGTC AATATATTAT CGTTCTGCAC CGGCGAATCAG CGGGGCATTC ACTTACGGG AGGCGAACCT GCGAAGTGCT GCCAGGTTTG TATCGACCT TATCGACCT TATCGACCT TATCGACGC AGTGTTTC TATCGACAGC CCATCGATGC CCATCGATGG GATGTTTCCG ATACAGCTAT CCGGCAAAAA GTGGACGAAT
40 45 50 55 60	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 851 901 1001 1151 1151 1251 1301 1451	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGATAC CCGGCCGAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGCG GCGAATATC CGGCAATATC CGGCAAAATT TCAAAAGCG GCGATTTCAT CCGCAAAAGT AAACGGCTTT CAAAGTTTTCAT CGCCAACAG AGAGCAGGAT ACGCCCAACAG AGAGCAGGAT ATCACGCCAA	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA TCTGTTTGTG AAGGCGATT CAAACCGATT CAAACCCA TTCTGTTGTG AAGCCGATTCGG CCGATAGCCA TTCGCGCCCG GCCGGCGGAAAACAA CAAAGATGGC TGATTGTCGA TTCGCGCCCG GCCGGCGGAAAACG ACGGAAAACG ACGGAAAACG CGATTTCGGC TGCATTCGGC TGCATATGGG AAGGGGACTT CGGCCCGGCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GAAAATGCCG TTGACGGC TTGACGGC TTGACAGCG TTGACAGCG TTGACAGCG TTGACAGCG TTGACAGCG TTGACAGCG CTTCATTTGC GCCGAGATGC GCGGCACGCA ACGCCACGCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTTCT AAGTGATGCA GGAATCAATC GCGATTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCC GTCAGCCTGA TCCGTCAGA TGGACGCAT TTCAAACG CCCGTCCAGA TGGACGGCT TTCAAAGCCG TGGCGGCGGG TGGCGGCGAAA ACTTCAACAC CCCTTCTTC	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGTT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA TAAATTTGTC AATATATTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT GCGAGGTTTG TTCGACGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC CCATCGATGC ATACAGCTAT CCGGCAAAAA GTGGACGAAT CAGCACCAAC
40 45 50 55 60	1 51 101 151 201 251 301 351 401 451 501 551 601 701 751 801 951 1001 1051 1151 1201 1251 1301 1351	ATGGCTAGCC CCCTGTTGTT CAGGTTCTCA GCGGCGGTTT CAAACCCAAA CCGGCCGAAAC CCGGCCGAAAC ACGATCCGTC ACAAATCAAG TTCAACCAAT ACGTGGCAA ACCCACTGTA AGTACAGCTA GGTTTGGTTG CTTTTATAAA GGTCGAGGCG GCGAATATC CGGAAAATT TCAAAAGGCG GCGATTTCAT CCGCAAAATT CAGCCAACAG GAAGTTTTCAT CCGCAACAG GAAGTTTTA AACGGCTTT GAAAGTTTTA AACGGCTTT CAAAGGCTTT CAAAGGCAGAT ATCACGCCAA AGAGCAGGAT ATCACGCCAA GTCGCCGAACAG AGAGCAGGAT ATCACGCCAA GTCGCCGCAT	TCTGAAAAAG AGGACAGGGC CGGAAGAAAA AATGAAGACG AGATAGTTTG ATATGGAAAA CAACCGGATA GGCAGGCGGG CCGAAAACAA TTCTGTTGTG AAGGCGATT CAAACCGATT CAAACCCA TTCTGTTGTG AAGGCGATT CAAACCCA GTCGCTTCCG TGATTGTCGA TTCGCGCCCG GCCGGCGAAAACA CGATTTCGCG CCCGGCGGAAAACG AAATGCTCGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC TGCATATCGGC AAGGGGACTT CGGCCCGGCC	AGACAGAGGC GCGCCATCCG TACAGGCAAT AGGGGGCGCA ACACCGAATAC CCAAGCACCG TGGCAAATAC GCAAACCGCC CGAATAGCGGC TTGTAGTGGC TTGAAAATT AAGAATGACG GCAGATGACG GCAGATGACG GCGGGAATAC GCGGCACGCA AGGGAATT TCGTATGCC GGCACGCA GCCGTCCGTC AGCAAAACT GGACAGAAC GGCGAAAAC GGACGGAAAA GGCGAAAAA GGCGAAGAAC GGCGGAAGAC GGCGGAAGAC GGCGGAAGAC GGCGGAAC GGCGGAACG GCCGTCCGTC AGCAAAAC GGACGGAAAA GGCCGAACGC GCCGTCCGTC AGCAAACC GGCCGACGCA GCCGTCCGTC AGCAAACC GGCCGACGCAACC GGCCGACGCACGC GCCATCGACC GACCGGTTCC	AAAGGAAGAT CACAAGGCGG GGCGGTGCGG AAATGATATG ACACCCGGC GATGCCGGG GGCGGACGGA GCAATACGGC GGTTCTCAAA TGGTGATTT CGTCGCAAAA AATAATTCT AAGTGATGCA GGAATCAATC GCGATTAGG CGCTGATTCC GTCAGCCTGA CCGGTATCC GTCAGCCTGA TCCGTGTTCA GTGACACG CCGTCCAGA TGGACGCAT TTCAAAGCCG TGGCGGCGG TGGCGGCGGG TGGCGGCGGAAA ATTTCAACAC GCCTACAAA ATTTCAACAC GTCGAGTTCG	GCGCCACAGG TCAAGATATG CAGCAACGGA CCGCAAAATG TTCGAATATG AATCGGAGCA ATGCAGGTT ATCCTGCCTC GGAAGGACGA TATAACGTTG TGGATGAAGA TAAATTTAT CGTTCTGCAC CGTCAATCAG CGGGGCATTC ACTTACGGGG AGGCGAACCT GCGAAGTGCT GCGAAGTTC GCAAGTTC GCAAGTTC GCAAGTTC TATCGACGC ACTACAGC CCATCGATGC ATACAGCTAT CCGGCAAAAA GTGGACGAAT CAGCACCAAC

-41-

	1601	GTTCGCAACA	CTTTACCGAC	CACCTGAAAT	CAGCCGACAT	CTTCGATGCC	
	1651	GCCCAATATC	CGGACATCCG	CTTTGTTTCC	ACCAAATTCA	ACTTCAACGG	
	1701	CAAAAAACTG	GTTTCCGTTG	ACGGCAACCT	GACCATGCAC	GGCAAAACCG	
~	1751	CCCCCGTCAA	ACTCAAAGCC	GAAAAATTCA	ACTGCTACCA	AAGCCCGATG	
5	1801	GCGAAAACCG	AAGTTTGCGG	CGGCGACTTC	AGCACCACCA	TCGACCGCAC	
	1851			TCGTTAACGT			
	1901	GCATCGACAT	CCAAATCGAG	GCAGCCAAAC	AATAAAAGCT	T	
	1	MASPDVKSAD	TLSKPAAPINI	SEKETEAKED	A POAGSOGOG	APSAQGGQDM	
10	51	AAVSEENTGN	GGAAATDKPK	NEDEGAONDM	PONAADTDSI.	TPNHTPASNM	
	101	PAGNMENOAP	DAGESEOPAN	QPDMANTADG	MOGDDPSAGG	ENAGNTAAOG	
	151	TNQAENNQTA	GSQNPASSTN	PSATNSGGDF	GRTNVGNSVV	IDGPSQNITL	
	201	THCKGDSCSG	NNFLDEEVQL	KSEFEKLSDA	DKISNYKKDG	KNDGKNDKFV	
	251	GLVADSVQMK	GINQYIIFYK	PKPTSFARFR	RSARSRRSLP	AEMPLIPVNQ	
15	301	ADTLIVDGEA	VSLTGHSGNI	FAPEGNYRYL	TYGAEKLPGG	SYALRVQGEP	
	351	SKGEMLAGTA	VYNGEVLHFH	TENGRPSPSR	GRFAAKVDFG	SKSVDGIIDS	
	401	GDGLHMGTQK	FKAAIDGNGF	KGTWTENGGG	DVSGKFYGPA	GEEVAGKYSY	
	451	RPTDAEKGGF	GVFAGKKEQD	GSGGGGATYK	VDEYHANARF	AIDHFNTSTN	
20	501	VGGFYGLTGS	VEFDQAKRDG	KIDITIPVAN	LQSGSQHFTD	HLKSADIFDA	
20	551 601			VSVDGNLTMH			
	601	AKTEVCGGDF	STTIDRIKWG	VDYLVNVGMT	KSVRIDIQIE	AAKQ*	
					,		
05	ΔG287NZ						
25	_1	ATGGCTAGCC	CCGATGTCAA	${\tt GTCGGCGGAC}$	ACGCTGTCAA	AACCTGCCGC	
	51	CCCTGTTGTT	TCTGAAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCCACAGG	
	101			GCGCCATCCG			
	151	GCGGCGGTTT	CGGAAGAAA	TACAGGCAAT	GGCGGTGCGG	CAGCAACGGA	
30	201 251	CCCCCCAAA	AATGAAGACG	AGGGGGCGCA ACACCGAATC	AAATGATATG	CCGCAAAATG	
50	301	CCGCCCGAIAC	AGAIAGIIIG	CCAAGCACCG	CAMCCCCCCCC	TTCGAATATG	
	351	GCCGGCAAAC	CAACCGGATA	TGGCAAATAC	GGCGGACGG	ARTCGGAGCA	
	401	ACGATCCGTC	GGCAGGCGGG	GAAAATGCCG	GCAATACGGC	TGCCCDAGGTG	
	451	ACAAATCAAG	CCGAAAACAA	TCAAACCGCC	GGTTCTCAAA	ATCCTGCCTC	
35	501	TTCAACCAAT	CCTAGCGCCA	CGAATAGCGG	TGGTGATTTT	GGAAGGACGA	
	551	ACGTGGGCAA	TTCTGTTGTG	ATTGACGGGC	CGTCGCAAAA	TATAACGTTG	
	601	ACCCACTGTA	AAGGCGATTC	${\tt TTGTAGTGGC}$	AATAATTTCT	TGGATGAAGA	
	651	AGTACAGCTA	AAATCAGAAT	${\tt TTGAAAAATT}$	AAGTGATGCA	GACAAAATAA	
40	701	GTAATTACAA	GAAAGATGGG	AAGAATGACG	GGAAGAATGA	TAAATTTGTC	
40	751	GGTTTGGTTG	CCGATAGTGT	GCAGATGAAG	GGAATCAATC	AATATATTAT	
	801 851	CTTTTTATAAA	CC'I'AAACCCA	CTTCATTTGC	GCGATTTAGG	CGTTCTGCAC	
	901	CCCCAMACCC	#CAMPORCE A	GCCGAGATGC TGGGGAAGCG	CGCTGATTCC	CGTCAATCAG	
	951	CGCCAATACGC	TGATTGTCGA	AAGGGAATTA	CCCCMAMCINC	CGGGGCATTC	
45	1001	CGGAAAAATT	GCCCGGCGGA	TCGTATGCCC	TCCGGTATCTG	ACTTACGGGG	
	1051	TCAAAAGGCG	AAATGCTCGC	GGGCACGGCA	CTCTACAACC	CCCA ACTCCT	
	1101			GCCGTCCGTC			
	1151	CCGCAAAAGT	CGATTTCGGC	AGCAAATCTG	TGGACGGCAT	TATCGACAGC	
	1201			TACGCAAAAA			
50	1251	AAACGGCTTT	AAGGGGACTT	GGACGGAAAA	TGGCGGCGGG	GATGTTTCCG	
	1301	GAAAGTTTTA	CGGCCCGGCC	GGCGAGGAAG	TGGCGGGAAA	ATACAGCTAT	
	1351	CGCCCAACAG	ATGCGGAAAA	GGGCGGATTC	GGCGTGTTTG	CCGGCAAAAA	
	1401	AGAGCAGGAT	GGATCCGGAG	GAGGAGGAGC	CACAAACGAC	GACGATGTTA	
55	1451	AAAAAGCTGC	CACTGTGGCC	ATTGCTGCTG	CCTACAACAA	TGGCCAAGAA	
33	1501	ATCAACGGTT	TCAAAGCTGG	AGAGACCATC	TACGACATTG	ATGAAGACGG	
	1551	CACAATTACC	AAAAAAGACG	CAACTGCAGC	CGATGTTGAA	GCCGACGACT	
	1601 1651	TIMAAGGICI	AACAAACC	AAAGTCGTGA	CTAACCTGAC	CAAAACCGTC	
	1701	AATACAAACA	TTA ACA ACCA	CGATGCCAAA AGTTAGCAGA	GRAAAAGCTG	CAGAATCTGA	
60	1751	ATACTGATGC	CCCTCTCCCA	GCAACCACCA	ACCCCMMCAA	GCTTTAGCAG	
	1801	GAAAATATAA	CGACATTTCC	TGAAGAGACT	DACACATAMA ACGCCTTGAA	TAMATIGGGA	
	1851	TGATGAAAAA	TTAGAAGCCG	TGGCTGATAC	CCTCCACAATA	CATCCCCVVC	
	1901	CATTCAACGA	TATCGCCGAT	TCATTGGATG	AAACCAACAAC	TAAGGCAGAC	
	1951	GAAGCCGTCA	AAACCGCCAA	TGAAGCCAAA	CAGACGGCCG	AAGAAACCAA	
65	2001	ACAAAACGTC	GATGCCAAAG	TAAAAGCTGC	AGAAACTGCA	GCAGGCAAAG	
	2051	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG	CAGCCGACAA	GGCCGAAGCT	
	2101	GTCGCTGCAA					

-42-

WO 01/64922 PCT/IB01/00452

	2151			ACAGTGCCGA		
	2201	CTGACAGCAA	ATTTGTCAGA	ATTGATGGTC	TGAACGCTAC	TACCGAAAAA
	2251	TTGGACACAC	GCTTGGCTTC	TGCTGAAAAA	TCCATTGCCG	ATCACGATAC
_	2301	TCGCCTGAAC	GGTTTGGATA	AAACAGTGTC	AGACCTGCGC	AAAGAAACCC
5	2351	GCCAAGGCCT	TGCAGAACAA	GCCGCGCTCT	CCGGTCTGTT	CCAACCTTAC
	2401	AACGTGGGTC	GGTTCAATGT	AACGGCTGCA	GTCGGCGGCT	ACAAATCCGA
	2451	ATCGGCAGTC	GCCATCGGTA	CCGGCTTCCG	CTTTACCGAA	AACTTTGCCG
	2501	CCAAAGCAGG	CGTGGCAGTC	GGCACTTCGT	CCGGTTCTTC	CGCAGCCTAC
	2551	CATGTCGGCG	TCAATTACGA	GTGGTAAAAG	CTT	
10						
	1	MASPDVKSAD	TLSKPAAPVV	SEKETEAKED	APQAGSQGQG	${\tt APSAQGGQDM}$
	51	AAVSEENTGN	GGAAATDKPK	NEDEGAQNDM	PQNAADTDSL	TPNHTPASNM
	101	PAGNMENQAP	DAGESEQPAN	QPDMANTADG	MQGDDPSAGG	ENAGNTAAQG
	151	TNQAENNQTA	GSQNPASSTN	PSATNSGGDF	GRTNVGNSVV	IDGPSQNITL
15	201	THCKGDSCSG	NNFLDEEVQL	KSEFEKLSDA	DKISNYKKDG	KNDGKNDKFV
	251	GLVADSVQMK	GINQYIIFYK	PKPTSFARFR	RSARSRRSLP	AEMPLIPVNQ
	301	ADTLIVDGEA	VSLTGHSGNI	FAPEGNYRYL	TYGAEKLPGG	SYALRVQGEP
	351	SKGEMLAGTA	VYNGEVLHFH	TENGRPSPSR	GRFAAKVDFG	SKSVDGIIDS
	401	GDGLHMGTQK	FKAAIDGNGF	KGTWTENGGG	DVSGKFYGPA	GEEVAGKYSY
20	451	RPTDAEKGGF	GVFAGKKEQD	GSGGGGATND	DDVKKAATVA	IAAAYNNGQE
	501	INGFKAGETI	YDIDEDGTIT	KKDATAADVE	ADDFKGLGLK	KVVTNLTKTV
	551	NENKQNVDAK	VKAAESEIEK	LTTKLADTDA	ALADTDAALD	ATTNALNKLG
	601	ENITTFAEET	KTNIVKIDEK	LEAVADTVDK	HAEAFNDIAD	SLDETNTKAD
	651	EAVKTANEAK	QTAEETKQNV	DAKVKAAETA	AGKAEAAAGT	ANTAADKAEA
25	701	VAAKVTDIKA	DIATNKDNIA	KKANSADVYT	REESDSKFVR	IDGLNATTEK
	751	LDTRLASAEK	SIADHDTRLN	GLDKTVSDLR	KETRQGLAEQ	AALSGLFQPY
	801	NVGRFNVTAA	VGGYKSESAV	AIGTGFRFTE	NFAAKAGVAV	GTSSGSSAAY
	851	HVGVNYEW*				

30 △G983 and hybrids

Bactericidal titres generated in response to $\Delta G983$ (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	NGH38	BZ133
Δ G 983	512	128	128

ΔG983 was also expressed as a hybrid, with ORF46.1, 741, 961 or 961c at its C-terminus:

	<u>Δ</u> G983-0	RF46.1				
35	1	ATGACTTCTG	CGCCCGACTT	CAATGCAGGC	GGTACCGGTA	TCGGCAGCAA
	51	CAGCAGAGCA	ACAACAGCGA	AATCAGCAGC	AGTATCTTAC	GCCGGTATCA
	101	AGAACGAAAT	GTGCAAAGAC	AGAAGCATGC	TCTGTGCCGG	TCGGGATGAC
	151	GTTGCGGTTA	CAGACAGGGA	TGCCAAAATC	AATGCCCCCC	CCCCGAATCT
	201	GCATACCGGA	GACTTTCCAA	ACCCAAATGA	CGCATACAAG	AATTTGATCA
40	251	ACCTCAAACC	$\mathtt{TGCAATTGAA}$	GCAGGCTATA	CAGGACGCGG	GGTAGAGGTA
	301	GGTATCGTCG	ACACAGGCGA	ATCCGTCGGC	AGCATATCCT	TTCCCGAACT
	351	GTATGGCAGA	AAAGAACACG	GCTATAACGA	AAATTACAAA	AACTATACGG
	401	CGTATATGCG	GAAGGAAGCG	CCTGAAGACG	GAGGCGGTAA	AGACATTGAA
	451	GCTTCTTTCG	ACGATGAGGC	CGTTATAGAG	ACTGAAGCAA	AGCCGACGGA
45	501	TATCCGCCAC	GTAAAAGAAA	TCGGACACAT	CGATTTGGTC	TCCCATATTA
	551	TTGGCGGGCG	TTCCGTGGAC	GGCAGACCTG	CAGGCGGTAT	TGCGCCCGAT
	601	GCGACGCTAC	ACATAATGAA	TACGAATGAT	GAAACCAAGA	ACGAAATGAT
	651	GGTTGCAGCC	ATCCGCAATG	CATGGGTCAA	GCTGGGCGAA	CGTGGCGTGC
	701	GCATCGTCAA	TAACAGTTTT	GGAACAACAT	CGAGGGCAGG	CACTGCCGAC
50	751	CTTTTCCAAA	TAGCCAATTC	GGAGGAGCAG	TACCGCCAAG	CGTTGCTCGA
	801	CTATTCCGGC	GGTGATAAAA	CAGACGAGGG	TATCCGCCTG	ATGCAACAGA
	851	GCGATTACGG	CAACCTGTCC	TACCACATCC	GTAATAAAAA	CATGCTTTTC
	901	ATCTTTTCGA	CAGGCAATGA	CGCACAAGCT	CAGCCCAACA	CATATGCCCT
	951	ATTGCCATTT	TATGAAAAAG	ACGCTCAAAA	AGGCATTATC	ACAGTCGCAG
55	1001	GCGTAGACCG	CAGTGGAGAA	AAGTTCAAAC	GGGAAATGTA	TGGAGAACCG
	1051	GGTACAGAAC	CGCTTGAGTA	TGGCTCCAAC	CATTGCGGAA	TTACTGCCAT
	1101	GTGGTGCCTG	TCGGCACCCT	ATGAAGCAAG	CGTCCGTTTC	ACCCGTACAA

WO 01/64922

-43-

PCT/IB01/00452

	1151	ACCCGATTCA	AATTGCCGGA	ACATCCTTTT	CCGCACCCAT	CGTAACCGGC
	1201	ACGGCGGCTC	${\tt TGCTGCTGCA}$	GAAATACCCG	TGGATGAGCA	ACGACAACCT
	1251	GCGTACCACG	${\tt TTGCTGACGA}$	CGGCTCAGGA	${\tt CATCGGTGCA}$	GTCGGCGTGG
_	1301	ACAGCAAGTT	${\tt CGGCTGGGGA}$	CTGCTGGATG	CGGGTAAGGC	CATGAACGGA
5	1351	CCCGCGTCCT	${\tt TTCCGTTCGG}$	CGACTTTACC	${\tt GCCGATACGA}$	AAGGTACATC
	1401	CGATATTGCC	TACTCCTTCC	GTAACGACAT	TTCAGGCACG	GGCGGCCTGA
	1451	TCAAAAAAGG	CGGCAGCCAA	CTGCAACTGC	ACGGCAACAA	CACCTATACG
	1501	GGCAAAACCA	${\tt TTATCGAAGG}$	CGGTTCGCTG	GTGTTGTACG	GCAACAACAA
	1551	ATCGGATATG	${\tt CGCGTCGAAA}$	CCAAAGGTGC	${\tt GCTGATTTAT}$	AACGGGGCGG
10	1601	CATCCGGCGG	CAGCCTGAAC	AGCGACGGCA	${\tt TTGTCTATCT}$	GGCAGATACC
	1651	GACCAATCCG	GCGCAAACGA	AACCGTACAC	ATCAAAGGCA	GTCTGCAGCT
	1701	GGACGGCAAA	${\tt GGTACGCTGT}$	ACACACGTTT	$\tt GGGCAAACTG$	CTGAAAGTGG
	1751	ACGGTACGGC	GATTATCGGC	GGCAAGCTGT	ACATGTCGGC	ACGCGGCAAG
	1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	${\tt CGTGTTCCCT}$	TCCTGAGTGC
15	1851			ATTCTTTCTT		
	1901			GACAGCGTCG		
	1951	GGCGACACGC	${\tt TGTCCTATTA}$	TGTCCGTCGC	GGCAATGCGG	CACGGACTGC
	2001	TTCGGCAGCG	GCACATTCCG	CGCCCGCCGG	TCTGAAACAC	GCCGTAGAAC
	2051	AGGGCGGCAG	${\tt CAATCTGGAA}$	AACCTGATGG	TCGAACTGGA	TGCCTCCGAA
20	2101	TCATCCGCAA	CACCCGAGAC	GGTTGAAACT	GCGGCAGCCG	ACCGCACAGA
	2151	TATGCCGGGC	ATCCGCCCCT	ACGGCGCAAC	TTTCCGCGCA	GCGGCAGCCG
	2201	TACAGCATGC	GAATGCCGCC	GACGGTGTAC	GCATCTTCAA	CAGTCTCGCC
	2251	GCTACCGTCT	ATGCCGACAG	TACCGCCGCC	${\tt CATGCCGATA}$	TGCAGGGACG
	2301	CCGCCTGAAA	GCCGTATCGG	ACGGGTTGGA	CCACAACGGC	ACGGGTCTGC
25	2351	GCGTCATCGC	GCAAACCCAA	CAGGACGGTG	${\tt GAACGTGGGA}$	ACAGGGCGGT
	2401	GTTGAAGGCA	AAATGCGCGG	CAGTACCCAA	ACCGTCGGCA	TTGCCGCGAA
	2451	AACCGGCGAA	AATACGACAG	CAGCCGCCAC	ACTGGGCATG	GGACGCAGCA
	2501	CATGGAGCGA	AAACAGTGCA	AATGCAAAAA	CCGACAGCAT	TAGTCTGTTT
20	2551	GCAGGCATAC	GGCACGATGC	GGGCGATATC	GGCTATCTCA	AAGGCCTGTT
30	2601	CTCCTACGGA	CGCTACAAAA	ACAGCATCAG	CCGCAGCACC	GGTGCGGACG
	2651			AACGGCACGC		
	2701			TGCCGCAACG		
	2751			TCAAACAGGA		
25	2801			AACAGCCTCA		
35	2851			GCAACCCTTG		
	2901			GCGACCTGAA		
	2951			ACTGCAGCAA		
	3001			GGTTGCCGGC		
40	3051			TGGCACGTTA		
40	3101			CGAGTCGGCG		
	3151		GCACTGGATC	CTCAGATTTG	CCAAACCAATT	
	3201			ATTTCGAACC	CGACGGGAAA	TACCACCTAT
	3251	TCGGCAGCAG	GGGGGAACTT	ATTTCGAACC GCCGAGCGCA	CGACGGGAAA GCGGCCATAT	TACCACCTAT CGGATTGGGA
15	3251 3301	TCGGCAGCAG AAAATACAAA	GGGGGAACTT GCCATCAGTT	ATTTCGAACC GCCGAGCGCA GGGCAACCTG	CGACGGGAAA GCGGCCATAT ATGATTCAAC	TACCACCTAT CGGATTGGGA AGGCGGCCAT
45	3251 3301 3351	TCGGCAGCAG AAAATACAAA TAAAGGAAAT	GGGGGAACTT GCCATCAGTT ATCGGCTACA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG
45	3251 3301 3351 3401	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC
45	3251 3301 3351 3401 3451	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA
45	3251 3301 3351 3401 3451 3501	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC
	3251 3301 3351 3401 3451 3501 3551	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT
45 50	3251 3301 3351 3401 3451 3501 3551 3601	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG
	3251 3301 3351 3401 3451 3501 3551 3601 3651	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG
	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCCGAGCT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG
	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701 3751	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCCGAGCT GCAGATATCG	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT
50	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701 3751 3801	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCCG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGG GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA
	3251 3301 3351 3401 3451 3551 3551 3651 3701 3751 3801 3851	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGC TAAGCGAAGG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCCGAGCT GCAGATATCG AGGCGATGCC ACGGCTTGGG	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGCCA TCTGCTTTCC
50	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701 3751 3801 3851 3901	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGC TAAGCGAAGG ACCGAAAACA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGATACA ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCCGAGCT GCAGATATCG AGGCGATGCC ACGCCTTGGG TTGGCCAGATA	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CCAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGCA TCTGCTTTCC TGGCGCAACT
50	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701 3751 3801 3851 3901 3951	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCA CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGCG TAAGCGAAGG ACCGAAAACA CAAAGACTAT	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCACCGA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ACGCATGCC ATGCTGGG TTGGCAGATA	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA
50	3251 3301 3351 3401 3451 3501 3551 3601 3651 3701 3851 3801 3851 3901 3951 4001	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGCG TAAGCGAAGC ACCGAAAACA CAAAGACTAT ATGCCGCACA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG TCGCCTCAA CGTTTCCACA CAAACGCCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCC GCCGCAGCAG AGGCCAGCAA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CCAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCACCGA GCCGTCAGCA GCCGTCAGCA GCCGTCAGCA GCCGTCAGCA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ACGGCTTGGG TTGGCAGATA TTGGCAGATA TTGGCAGATA TTGGCAGATA ATGGCAGATC ATATCTTTAT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA CGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCATC
50 55	3251 3301 3351 3401 3451 3501 3551 3601 3751 37701 3751 3801 3851 3901 3951 4001 4051	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGCG TAAGCGAAGC ACCGAAAACA CAAAGACTAT ATGCCGCACA CCCATCAAAG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG TCCGCCTCAA CGTTTCCACA CAAACGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG AGGCATAGAA GGATTGGAGC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACG GATATATACA CCTGACCGAC ATGCCGTAG ACCCGATACA CCAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCACCGA GCCGTCAGCA TCTTCGGCGA TCTTCGGCGA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG GCAGATATCG ACGCGATGCC ACGGCTTGGG TTGGCAGATA TTGGCAGATA TTGGCAGATA ATATCTTTAT AAATACGCT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA CGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCAT TGGGCGCAT TGGGCGCAT
50	3251 3301 3351 3401 3451 3501 3551 3601 3751 3751 3801 3851 3901 3951 4001 4051 4101	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGCG TAAGCGAAG ACCGAAAACA CAAAGACTAT ATGCCGCACA CCCATCAAAG CACGGCACA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG TCCGCCTCAA CGTTTCCACA CAAACGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG AGGCATAGAA GGATTGGAGC CCTATCAACC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACG GATATATACA CCTGACCGAC ATGCCGTAG ACCCGATACA CCAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCACCGA GCCGTCAGCA TGTTCGGCGA GCCGTCAGCA TGTTCGGCGA GCTCTCGCGA GCTCTCAGCA TGTTCGGGGA GCTCCGCGA	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG GCAGATATCG ACGCGATGCC ATGGCAGATA TTGGCAGATA TTGGCAGATA TTGGCAGATA TTGGCAGATA ATACCGCT ATATCTTTAT AAATACGGCT GGGCGCGATC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA CGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCATC TGGGCGCAT CGAGGCAT CGAGGCAT CGAGGCCAT CGAGGCCAT GCATTGCCGA
50 55	3251 3301 3351 3401 3451 3501 3551 3601 3701 3751 3801 3851 3901 3951 4001 4051 4101 4151	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGC TAAGCGAAGC ACCGAAAACA CAAAGACTAT ATGCCGCACA CCCATCAAAG CACGGCACAT AAGGGAAATC	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG AGGCATAGAA GGATTGGAGC CCTATCAAGC CCCCTCAGC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCAGCAC TCGTCAGCA TCGTCAGCA GCCGTCAGCA TGTTCGGGGA GCGTCAGCA TGTTCGGGGA GGTCGCAGAT GACAATTTTG	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATC AGGCGATGCC AGGCGTTGGG TTGGCAGAT TTGGGCAGT TTGGGCAGT CATATCTTAT AAATACGCT GGGCGCGATC CCGATGCGC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA CGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCATC TGGGCGCAT TGGGCGCAT TGGGCGCAT GCATTGCCGA ATACGCCAAA
50 55	3251 3301 3351 3401 3551 3501 3551 3601 3751 3851 3851 3901 3951 4001 4051 4101 4201	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCA CGGACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGC TAAGCGAAGA ACCGAAAACA CAAAGACTAT ATGCCGCACA CCCATCAAAG CACGGCACAT AAGGGAAATC TACCCGTCCC	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG AGGCATAGAA GGATTGGAGC CCTATCAAGC CCTATCAAGC CCTTATCAAGC CCTTATCAAGC CCCCTTACCATT	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCAGCGA GCCGTCAGCA TGTTCGGGGA GCCGTCAGCA TGTTCGGGGA GCGTCAGCA TGTTCGGGGA GCGTCAGCA TGTTCGGGGA GCGTCAGCA TGTTCGGGGA GCGTCAGCA TGTTCGGGGA GCGTCAGAT GACAATTTTG CCGAAATATC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ATGGCTTGGG TTGGGCAGTT ATATCTTTAT AAATACGGCT GGGCGCGATC CCGATGCG CCGATGCG CCGATGCGC CCGATGCGC CCGATGCGC CCGATGCGC CCGATGCGC CCGTTCAAACT	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCA GGCAGCCATC TGGGCGCAT TGGGCGCAT GCATTGCCGA ATACGCCAAA TGGAGCAGCG
50 55	3251 3301 3351 3401 3451 3501 3551 3601 3751 3851 3851 3951 4001 4051 4101 4201 4251	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG CATCGGCGC TAAGCGAAGC ACCGAAAACA CCAAAGACTAT ATGCCGCACA CCCATCAAAG CACGGCACAT AAGGGAAATC TACCCGTCCC TTACGGCCAAA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGAGG TCCGCCTCAA CGTTTCCACA CAAACGCGCC CGAAGCCTT GCAGGAAAA CTCAAACATT AGATGGCGCG GCCGCAGCAG AGGCATAGAA GGATTGGAGC CCTATCAAGC CCGCGTCAGC CTTACCATT GAAAACATCA	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGTAG ACCCGATACA CAACGGACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCACGAC TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTTCGGGAAT CCACAATTTTG CCGAAATATC CCTCCTCAAC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ATGGCTAGGAT ATGCTGACG TTGGGCAGTC ATGCTGACG CGCTTGGG CGCTTGGG CGCTTGGC CGCCGATC CGGCGCGCC CGTTCAAACT CGTGCCGCC	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGGATA CGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGCA TCTGCTTTCC CAAAACCCA GGCAGCATC TGGCCGCAT TGGGCGCAT CGACACCA TGGCCGCAT CAAAACCCA TGGCCGCAT TGGGCGCAT TGGGCGCAT TGGGCGCAT TGGGCGCAT TGGACCAAA TGGAGCAGCG TCAAACCCA
505560	3251 3301 3351 3401 3451 3501 3551 3601 3751 3851 3901 3851 3901 4001 4051 4101 4201 4251 4301	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCG CGAACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGATT GGCAATGCCG TAAGCGAAACA ACCGAAACA ATGCCGCACA CCCATCAAAG CCATCAAAG CACGGCACAT ATGCCGCACA CACGGCACAT AAGGGAAATC TACCCGTCCC TTACGGCAAA AAAATGTCAA	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG AGGCGCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG AGGCATAGAA GGATTGGAGC CCTATCAAGC CCGCCGTCAGC CTTACCATT GAAAACATC ACTGGCAGAC ACTGGCAGAC ACTGGCAGAC CTTACCATC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAG ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCCGCGA GCCGTCAGCA TGTTCGGGGG GCTCTCAGCA TGTTCGGGAGAT GGCCACATTTTG CCGAAATATC CCGAAATATC CCTCCTCAAC CAACGCCACC	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ACGGCTTGGG TTGGCAGATA TTGGGCAGTC ATATCTTTAT AAATACGGCT CGGATGCC CCGATGCGC CGTTCAAACT CGTGCCGCCG CGAAGACAGG	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCAT TGGGCGCAT TGGGCGCAT TGGGCGCAT TGGAGCACAT TGGAGCACAC ATACGCCAA TGGAGCAGCA TCGAAACCCCA CGTACCGTTT
50 55	3251 3301 3351 3401 3451 3501 3551 3601 3751 3851 3851 3951 4001 4051 4101 4201 4251	TCGGCAGCAG AAAATACAAA TAAAGGAAAT TCCATTCCCC GGTAGTCCCA GCGACACCAT CCGCTCCCAA GCCCAAAATA GCTTGCCGAC GCGACGGAT GGCAATGCCG CATCGGCGCG TAAGCGAAACA CAAAGACTAT ATGCCGCACA CCCATCAAAG CCATCAAAG CCACATCAAAG TACCCGTCCC TACCGGCACAT ATGCCGCACA TACGCGCACAT ATGCCGCACA TACGCGCACAT AAGGGAAATC TACCCGTCCC TTACGGCAAA AAAATGTCAA GACGGTAAAG	GGGGGAACTT GCCATCAGTT ATCGGCTACA CTTCGACAAC TTGACGGATT CCCGCCGACG AGGCGCGACG AGGCGCTCAA CGTTTCCACA CAAACGCGCC CCGAAGCCTT GCAGGAGAAA CTCAAACATT AGATGGCGCG AGGCATAGAA GGATTGGAGC CCTATCAAGC CCGCCGTCAGC CTTACCATT GAAAACATC ACTGGCAGAC ACTGGCAGAC ACTGGCAGAC CTTACCATC	ATTTCGAACC GCCGAGCGCA GGGCAACCTG TTGTCCGCTT CATGCCTCAC TAGCCTTTAC GCTATGACGG GATATATACA CCTGACCGAC ATGCCGGTAC ACCCGATACA CAACGGCACT TTGTCGGCGC GCTGTCATGC CATCAACGAT CCATCCGCGA GCCGTCAGCA TGTCGGGGA GCTCTCAGCA TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTTCGGGGA TGTCGCAAATATC CCGAAATATC CCTCCTCAAC CAACGCCACC TTTTGAGAAG	CGACGGGAAA GCGGCCATAT ATGATTCAAC TTCCGATCAC ATTCCGATTC CGCATCCATT GCCACAGGC GCTACGACAT AACCGCAGCA TATGCTGACG GCCCGAGCT GCAGATATCG AGGCGATGCC ACGGCTTGGG TTGGCAGATA TTGGGCAGTC ATATCTTTAT AAATACGGCT CGGATGCC CCGATGCGC CGTTCAAACT CGTGCCGCCG CGAAGACAGG	TACCACCTAT CGGATTGGGA AGGCGGCCAT GGGCACGAAG TGATGAAGCC GGGACGATA GGCGGCTATC AAAAGGCGTT CCGGACAACG CAAGGAGTAG GGACAGATCG TTAAAAACAT GTGCAGGGCA TCTGCTTTCC TGGCGCAACT CAAAACCCCA GGCAGCCAT TGGGCGCAT TGGGCGCAT TGGGCGCAT TGGAGCACAT TGGAGCACAC ATACGCCAA TGGAGCAGCA TCGAAACCCCA CGTACCGTTT

WO 01/64922 PCT/IB01/00452 -44-

MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMLCAGRDD 51 VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV 101 GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE 151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGIAPD 5 ATLHIMNTND ETKNEMMVAA IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD 201 LFOIANSEEO YROALLDYSG GDKTDEGIRL MOOSDYGNLS YHIRNKNMLF IFSTGNDAQA QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP 301 351 GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG 401 TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG 10 PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT 451 GKTIIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK 551 601 GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE 651 GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEOGGSNLE NLMVELDASE 15 SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA 701 ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG 751 VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF 801 851 AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL 901 GGVNVPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG 20 LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGGFTGA TAATGKTGAR 951 1001 NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLD GGGGTGSSDL ANDSFIRQVL DRQHFEPDGK YHLFGSRGEL AERSGHIGLG 1051 1101 KIQSHQLGNL MIQQAAIKGN IGYIVRFSDH GHEVHSPFDN HASHSDSDEA 1151 GSPVDGFSLY RIHWDGYEHH PADGYDGPQG GGYPAPKGAR DIYSYDIKGV 25 AQNIRLNLTD NRSTGQRLAD RFHNAGSMLT QGVGDGFKRA TRYSPELDRS 1201 GNAAEAFNGT ADIVKNIIGA AGEIVGAGDA VQGISEGSNI AVMHGLGLLS 1251 TENKMARIND LADMAQLKDY AAAAIRDWAV QNPNAAQGIE AVSNIFMAAI 1301 1351 PIKGIGAVRG KYGLGGITAH PIKRSQMGAI ALPKGKSAVS DNFADAAYAK 1401 YPSPYHSRNI RSNLEQRYGK ENITSSTVPP SNGKNVKLAD QRHPKTGVPF 30 DGKGFPNFEK HVKYDTLEHH HHHH* 1451 ΔG983-741 1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA 35 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA 51 AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC 151 GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT 201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA 251 40 301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT 351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA 401 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA 451 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA 501 45 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT 551 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC 651 GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC 701 751 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA 50 801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA 851 GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAA CATGCTTTTC 901 ATCTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT 951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG 1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG 55 1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA 1101 ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC 1151 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT 1201 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG 1251 60 1301 ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC 1351 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA 1401 1451 TCAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG 1501 GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA 65 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG 1551 1601 CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC

GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT

1651

	1701	GGACGGCAAA	GGTACGCTGT	ACACACGTTT	GGGCAAACTG	CTGAAAGTGG
	1751				ACATGTCGGC	
	1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	CGTGTTCCCT	TCCTGAGTGC
_	1851	CGCCAAAATC	GGGCAGGATT	ATTCTTTCTT	CACAAACATC	GAAACCGACG
5	1901				AAAAAACAGC	
	1951				GGCAATGCGG	
	2001				TCTGAAACAC	
	2051				TCGAACTGGA	
10	2101 2151				GCGGCAGCCG TTTCCGCGCA	
10	2201				GCATCTTCAA	
	2251				CATGCCGATA	
	2301				CCACAACGGC	
	2351				GAACGTGGGA	
15	2401				ACCGTCGGCA	
	2451	AACCGGCGAA	AATACGACAG	CAGCCGCCAC	ACTGGGCATG	GGACGCAGCA
	2501	CATGGAGCGA	${\tt AAACAGTGCA}$	AATGCAAAAA	CCGACAGCAT	TAGTCTGTTT
	2551				GGCTATCTCA	
20	2601				CCGCAGCACC	
20	2651				TGATGCAGCT	
	2701				GGAGATTTGA	
	2751				TGCATTCGCC	
	2801 2851				CTGAAGGCAC AGCGATAAAG	
25	2901				CGGACGCGAC	
23	2951				CCGGCAAGAC	
	3001				CTGGGCGCGG	
	3051				CAGCTACGCC	
	3101				TAGGCTACCG	
30	3151	GGATCCGGAG	GGGGTGGTGT	CGCCGCCGAC	ATCGGTGCGG	GGCTTGCCGA
	3201	TGCACTAACC	GCACCGCTCG	ACCATAAAGA	CAAAGGTTTG	CAGTCTTTGA
	3251				AACTGAAGCT	
	3301				AGCCTCAATA	
35	3351				TATCCGCCAA	
33	3401				AGTTCCAAGT	
	3451 3501				GAGCAAATAC GTTCAGAATC	
	3551				CCGAAGGCGG	
	3601				GCCGGCGGAA	
40	3651				CGGCAAAATC	
	3701				CCGCCGATAT	
	3751	GGAAAACGCC	ATGCCGTCAT	CAGCGGTTCC	GTCCTTTACA	ACCAAGCCGA
	3801	GAAAGGCAGT	TACTCCCTCG	GTATCTTTGG	CGGAAAAGCC	CAGGAAGTTG
. ~	3851				GCATACGCCA	TATCGGCCTT
45	3901	GCCGCCAAGC	AACTCGAGCA	CCACCACCAC	CACCACTGA	
	1				AGIKNEMCKD	
	51 101				NLINLKPAIE NYTAYMRKEA	
50	151				SHIIGGRSVD	
50	201				RGVRIVNNSF	
	251				MQQSDYGNLS	
	301				TVAGVDRSGE	
	351	GTEPLEYGSN	HCGITAMWCL	SAPYEASVRF	TRTNPIQIAG	TSFSAPIVTG
55	401	TAALLLQKYP	WMSNDNLRTT	LLTTAQDIGA	VGVDSKFGWG	LLDAGKAMNG
	451	PASFPFGDFT	ADTKGTSDIA	YSFRNDISGT	GGLIKKGGSQ	LQLHGNNTYT
	501				NGAASGGSLN	
	551				LKVDGTAIIG	
60	601			~	ETDGGLLASL	
60	651 701				AVEQGGSNLE	
	701 751				AAAVQHANAA	
	801				TGLRVIAQTQ GRSTWSENSA	
	851				GADEHAEGSV	
65	901				EKGSALGWSG	
- -	951			_	YTVTGGFTGA	
	1001				GSKQYGNHSG	

-46-

GSGGGGVAAD IGAGLADALT APLDHKDKGL QSLTLDQSVR KNEKLKLAAQ 1051 GAEKTYGNGD SLNTGKLKND KVSRFDFIRO IEVDGQLITL ESGEFQVYKO SHSALTAFQT EQIQDSEHSG KMVAKRQFRI GDIAGEHTSF DKLPEGGRAT 1151 1201 YRGTAFGSDD AGGKLTYTID FAAKQGNGKI EHLKSPELNV DLAAADIKPD 5 GKRHAVISGS VLYNQAEKGS YSLGIFGGKA QEVAGSAEVK TVNGIRHIGL 1251 1301 AAKOLEHHHH HH* ∆G983-961 10 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA 1. CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA 51 AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC 151 GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA 201 15 251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT 301 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA 401 451 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA 20 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA 501 TTGGCGGCCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT 551 601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT 651 GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC 701 GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGCCAGG CACTGCCGAC 25 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA 751 801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC 851 ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT 901 951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG 30 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG 1001 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT 1051 1101 GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC 1151 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT 1201 35 1251 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG 1301 ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC 1351 1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG 1451 40 GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA 1501 1551 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC 1601 1651 GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT 1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG 45 1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG 1801 GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG 1851 1901 GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA GGCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC 1951 50 2001 TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC AGGGCGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA 2051 2101 TCATCCGCAA CACCCGAGAC GGTTGAAACT GCGGCAGCCG ACCGCACAGA TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG 2151 2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC 55 GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG 2251 CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC 2301 2351 GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA 2401 AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA 2451 60 CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT 2501 GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT 2551 2601 CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG 2651 2701 GGCGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCGAAGG 65 2751 CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTCGCC GAAAAAGGCA GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA 2801 2851 CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCCTGTT

-47-

	2901	TGCAACGGCG	GGCGTGGAAC	GCGACCTGAA	CGGACGCGAC	TACACGGTAA
	2951	CGGGCGGCTT	TACCGGCGCG	ACTGCAGCAA	CCGGCAAGAC	GGGGGCACGC
	3001	AATATGCCGC	ACACCCGTCT	GGTTGCCGGC	CTGGGCGCGG	ATGTCGAATT
_	3051	CGGCAACGGC	TGGAACGGCT	TGGCACGTTA	CAGCTACGCC	GGTTCCAAAC
5	3101	AGTACGGCAA	CCACAGCGGA	CGAGTCGGCG	TAGGCTACCG	GTTCCTCGAG
	3151			CGCCACAAAC		
	3201			CTGCCTACAA		
	3251			ATCTACGACA		
10	3301			AGCCGATGTT		
10	3351			TGACTAACCT		
	3401			AAAGTAAAAG		
	3451			AGACACTGAT		
	3501			CCAACGCCTT		
15	3551			ACTAAGACAA		
13	3601 3651			TACCGTCGAC ATGAAACCAA		
	3701			AAACAGACGG		
	3751			TGCAGAAACT		
	3801			CTGCAGCCGA		
20	3851			GCTGATATCG		
20	3901			CGACGTGTAC		
	3951			GTCTGAACGC		
	4001			AAATCCATTG		
	4051			GTCAGACCTG		
25	4101			TCTCCGGTCT		
	4151			GCAGTCGGCG		
	4201	GTCGCCATCG	GTACCGGCTT	CCGCTTTACC	GAAAACTTTG	CCGCCAAAGC
	4251	AGGCGTGGCA	GTCGGCACTT	CGTCCGGTTC	TTCCGCAGCC	TACCATGTCG
	4301	GCGTCAATTA	CGAGTGGCTC	GAGCACCACC	ACCACCACCA	CTGA
30						
	1	MTSAPDFNAG	GTGIGSNSRA	TTAKSAAVSY	AGIKNEMCKD	RSMLCAGRDD
	51			DFPNPNDAYK		
	101			KEHGYNENYK		
25	151			VKEIGHIDLV		
35	201			IRNAWVKLGE		
	251			GDKTDEGIRL		
	301		-	YEKDAQKGII		
	351			SAPYEASVRF		
40	401			LLTTAQDIGA		
40	451			YSFRNDISGT RVETKGALIY		_
	501 551			GTLYTRLGKL		
	601			GODYSFFTNI		
	651			AHSAPAGLKH		
45	701			IRPYGATERA		
	751			AVSDGLDHNG	~	
	801			NTTAAATLGM	~ ~ ~	
	851			RYKNSISRST		
	901			YDLLKQDAFA		_
50	951			GVERDLNGRD		
	1001	NMPHTRLVAG	LGADVEFGNG	WNGLARYSYA	GSKQYGNHSG	RVGVGYRFLE
	1051	GGGGTGSATN	DDDVKKAATV	AIAAAYNNGQ	EINGFKAGET	IYDIDEDGTI
	1101	TKKDATAADV	EADDFKGLGL	KKVVTNLTKT	VNENKQNVDA	KVKAAESEIE
	1151	KLTTKLADTD	AALADTDAAL	DATTNALNKL	GENITTFAEE	TKTNIVKIDE
55	1201	KLEAVADTVD	KHAEAFNDIA	DSLDETNTKA	DEAVKTANEA	KQTAEETKQN
	1251	VDAKVKAAET	AAGKAEAAAG	TANTAADKAE	AVAAKVTDIK	ADIATNKDNI
	1301	AKKANSADVY	TREESDSKFV	RIDGLNATTE	KLDTRLASAE	KSIADHDTRL
	1351	NGLDKTVSDL	RKETRQGLAE	QAALSGLFQP	YNVGRFNVTA	AVGGYKSESA
CO	1401	VAIGTGFRFT	ENFAAKAGVA	VGTSSGSSAA	YHVGVNYEWL	ЕНННННН*
60						
	<u>ΔG983-9</u>					
	1			CAATGCAGGC		
65	51			AATCAGCAGC		
O.J	101			AGAAGCATGC		
	151			TGCCAAAATC		
	201	GCATACCGGA	GACTITCCAA	ACCCAAATGA	CGCATACAAG	AATTTGATCA

	251				CAGGACGCGG	
	301				AGCATATCCT	
	351				AAATTACAAA	
5	401				GAGGCGGTAA	
3	451				ACTGAAGCAA	
	501				CGATTTGGTC	
	551				CAGGCGGTAT	
	601				GAAACCAAGA	
10	651				GCTGGGCGAA	
10	701				CGAGGGCAGG	
	751				TACCGCCAAG TATCCGCCTG	
	801 851				GTAATAAAAA	
	901				CAGCCCAACA	
15	951				AGGCATTATC	
15	1001				GGGAAATGTA	
	1051				CATTGCGGAA	
	1101				CGTCCGTTTC	
	1151				CCGCACCCAT	
20	1201				TGGATGAGCA	
	1251				CATCGGTGCA	
	1301				CGGGTAAGGC	
	1351				GCCGATACGA	
	1401				TTCAGGCACG	
25	1451				ACGGCAACAA	
	1501				GTGTTGTACG	
	1551				GCTGATTTAT	
	1.601				TTGTCTATCT	
	1651	GACCAATCCG	GCGCAAACGA	AACCGTACAC	ATCAAAGGCA	GTCTGCAGCT
30	1701	GGACGGCAAA	GGTACGCTGT	ACACACGTTT	GGGCAAACTG	CTGAAAGTGG
	1751	ACGGTACGGC	GATTATCGGC	GGCAAGCTGT	ACATGTCGGC	ACGCGGCAAG
	1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	CGTGTTCCCT	TCCTGAGTGC
	1851	CGCCAAAATC	GGGCAGGATT	ATTCTTTCTT	CACAAACATC	GAAACCGACG
2.4	1901				AAAAAACAGC	
35	1951				GGCAATGCGG	
	2001	TTCGGCAGCG	GCACATTCCG	CGCCCGCCGG	TCTGAAACAC	GCCGTAGAAC
	2051	AGGGCGGCAG	CAATCTGGAA	AACCTGATGG	TCGAACTGGA	TGCCTCCGAA
	2101				GCGGCAGCCG	
40	2151				TTTCCGCGCA	
40	2201				GCATCTTCAA	
	2251				CATGCCGATA	
	2301				CCACAACGGC	
	2351				GAACGTGGGA	
45	2401				ACCGTCGGCA	
43	2451				ACTGGGCATG	
	2501 2551				CCGACAGCAT	
					GGCTATCTCA CCGCAGCACC	
	2601 2651				TGATGCAGCT	
50	2701				GGAGATTTGA	
50	2751				TGCATTCGCC	
	2801				CTGAAGGCAC	
	2851				AGCGATAAAG	
	2901				CGGACGCGAC	
55	2951				CCGGCAAGAC	
	3001				CTGGGCGCGG	
	3051				CAGCTACGCC	
	3101				TAGGCTACCG	
	3151				GACGACGATG	
60	3201	TGCCACTGTG	GCCATTGCTG	CTGCCTACAA	CAATGGCCAA	GAAATCAACG
	3251	GTTTCAAAGC	TGGAGAGACC	ATCTACGACA	TTGATGAAGA	CGGCACAATT
	3301	ACCAAAAAAG	ACGCAACTGC	AGCCGATGTT	GAAGCCGACG	ACTTTAAAGG
	3351				GACCAAAACC	
	3401	ACAAACAAAA	CGTCGATGCC	AAAGTAAAAG	CTGCAGAATC	TGAAATAGAA
65	3451	AAGTTAACAA	CCAAGTTAGC	AGACACTGAT	GCCGCTTTAG	CAGATACTGA
	3501	TGCCGCTCTG	GATGCAACCA	CCAACGCCTT	GAATAAATTG	GGAGAAAATA
	3551	TAACGACATT	TGCTGAAGAG	ACTAAGACAA	ATATCGTAAA	AATTGATGAA

-49-

	3601	AAATTAGAAG	CCGTGGCTGA	TACCGTCGAC	AAGCATGCCG	AAGCATTCAA
	3651	CGATATCGCC	GATTCATTGG	ATGAAACCAA	CACTAAGGCA	GACGAAGCCG
	3701	TCAAAACCGC	${\tt CAATGAAGCC}$	AAACAGACGG	CCGAAGAAAC	CAAACAAAAC
_	3751	GTCGATGCCA	AAGTAAAAGC	TGCAGAAACT	GCAGCAGGCA	AAGCCGAAGC
5	3801	TGCCGCTGGC	ACAGCTAATA	CTGCAGCCGA	CAAGGCCGAA	GCTGTCGCTG
	3851	CAAAAGTTAC	CGACATCAAA	GCTGATATCG	CTACGAACAA	AGATAATATT
	3901	GCTAAAAAAG	CAAACAGTGC	CGACGTGTAC	ACCAGAGAAG	AGTCTGACAG
	3951	CAAATTTGTC	AGAATTGATG	GTCTGAACGC	TACTACCGAA	AAATTGGACA
	4001	CACGCTTGGC	TTCTGCTGAA	AAATCCATTG	CCGATCACGA	TACTCGCCTG
10	4051	AACGGTTTGG	ATAAAACAGT	GTCAGACCTG	CGCAAAGAAA	CCCGCCAAGG
	4101	CCTTGCAGAA	CAAGCCGCGC	TCTCCGGTCT	GTTCCAACCT	TACAACGTGG
	4151	GTCTCGAGCA	CCACCACCAC	CACCACTGA		
	1			TTAKSAAVSY		
15	51			DFPNPNDAYK		
	101			KEHGYNENYK		
	151			VKEIGHIDLV		
	201			IRNAWVKLGE		
••	251		~	GDKTDEGIRL	~~	
20	301	~	~	YEKDAQKGII		
	351			SAPYEASVRF	~	
	401			LLTTAQDIGA		
	451			YSFRNDISGT		
25	501			RVETKGALIY		
25	551			GTLYTRLGKL		
	601			GQDYSFFTNI		
	651			AHSAPAGLKH	~	
	701			IRPYGATFRA		
20	751		~	AVSDGLDHNG		
30	801	~		NTTAAATLGM		
	851			RYKNSISRST		
	901			YDLLKQDAFA		
	951	~		GVERDLNGRD		
2.5	1001			WNGLARYSYA		
35	1051			AIAAAYNNGQ		
	1101			KKVVTNLTKT	~	
	1151			DATTNALNKL		
	1201			DSLDETNTKA		~ ~
40	1251			TANTAADKAE		
40	1301			RIDGLNATTE		
	1351	NGLDKTVSDL	RKETRQGLAE	QAALSGLFQP	YNVGLEHHHH	HH*

△G741 and hybrids

Bactericidal titres generated in response to $\Delta G741$ (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	MC58	NGH38	F6124	BZ133
∆G741	512	131072	>2048	16384	>2048

As can be seen, the $\Delta G741$ -induced anti-bactericidal titre is particularly high against heterologous strain MC58.

ΔG741 was also fused directly in-frame upstream of proteins 961, 961c, 983 and ORF46.1:

<u>∆G741-961</u>

	1	ATGGTCGCCG	CCGACATCGG	TGCGGGGCTT	GCCGATGCAC	TAACCGCACC
50	51	GCTCGACCAT	AAAGACAAAG	GTTTGCAGTC	TTTGACGCTG	GATCAGTCCG
	101	TCAGGAAAAA	CGAGAAACTG	AAGCTGGCGG	CACAAGGTGC	GGAAAAAACT
	151	TATGGAAACG	GTGACAGCCT	CAATACGGGC	AAATTGAAGA	ACGACAAGGT
	201	CAGCCGTTTC	GACTTTATCC	GCCAAATCGA	AGTGGACGGG	CAGCTCATTA

WO 01/64922

	251	CCTTGGAGAG	TGGAGAGTTC	CAAGTATACA	AACAAAGCCA	TTCCGCCTTA
	301	ACCGCCTTTC	AGACCGAGCA	AATACAAGAT	TCGGAGCATT	CCGGGAAGAT
	351	GGTTGCGAAA	CGCCAGTTCA	GAATCGGCGA	CATAGCGGGC	GAACATACAT
	401	CTTTTGACAA	GCTTCCCGAA	GGCGGCAGGG	CGACATATCG	CGGGACGGCG
5	451	TTCGGTTCAG	ACGATGCCGG	CGGAAAACTG	ACCTACACCA	TAGATTTCGC
	501	CGCCAAGCAG	GGAAACGGCA	AAATCGAACA	TTTGAAATCG	CCAGAACTCA
	551	ATGTCGACCT	GGCCGCCGCC	GATATCAAGC	CGGATGGAAA	ACGCCATGCC
	601	GTCATCAGCG	GTTCCGTCCT	TTACAACCAA	GCCGAGAAAG	GCAGTTACTC
	651			AAGCCCAGGA		
10	701			CGCCATATCG		
	751			ATCCGCCACA		
	801			CTGCTGCCTA		
	851			ACCATCTACG		
	901			TGCAGCCGAT		
15	951			TCGTGACTAA		
10	1001			GCCAAAGTAA		
	1051	-		AGCAGACACT		
	1101			CCACCAACGC		
20	1151			GAGACTAAGA		
20	1201			TGATACCGTC TGGATGAAAC		
	1251					
	1301			GCCAAACAGA		
	1351			AGCTGCAGAA		
25	1401			ATACTGCAGC		
25	1451			AAAGCTGATA		
	1501			TGCCGACGTG		
	1551			ATGGTCTGAA		
	1601			GAAAAATCCA		
20	1651			AGTGTCAGAC		
30	1701			CGCTCTCCGG		
	1751			GCTGCAGTCG		
	1801			CTTCCGCTTT		
	1851			CTTCGTCCGG		
0.5	1901	TCGGCGTCAA	TTACGAGTGG	CTCGAGCACC	ACCACCACCA	CCACTGA
35						
	1			KDKGLQSLTL		
	51			DFIRQIEVDG	-	
	101	~ ~ ~		RQFRIGDIAG		
40	151		~	GNGKIEHLKS		
40	201			FGGKAQEVAG		
	251			VAIAAAYNNG	**	
	301			LKKVVTNLTK		
	351			LDATTNALNK		
	401			ADSLDETNTK		
45	451	NVDAKVKAAE	TAAGKAEAAA	GTANTAADKA	EAVAAKVTDI	KADIATNKDN
	501			VRIDGLNATT		
	551	LNGLDKTVSD	LRKETRQGLA	EQAALSGLFQ	PYNVGRFNVT	AAVGGYKSES
	601	AVAIGTGFRF	TENFAAKAGV	AVGTSSGSSA	AYHVGVNYEW	ГЕННИННН
~ o						
50						
	<u>ΔG741-9</u>					
	1	ATGGTCGCCG	CCGACATCGG	TGCGGGGCTT	GCCGATGCAC	TAACCGCACC
	51	GCTCGACCAT	AAAGACAAAG	GTTTGCAGTC	TTTGACGCTG	GATCAGTCCG
	101	TCAGGAAAAA	CGAGAAACTG	AAGCTGGCGG	CACAAGGTGC	GGAAAAAACT
55	151	TATGGAAACG	GTGACAGCCT	CAATACGGGC	AAATTGAAGA	ACGACAAGGT
	201	CAGCCGTTTC	GACTTTATCC	GCCAAATCGA	AGTGGACGGG	CAGCTCATTA
	251	CCTTGGAGAG	TGGAGAGTTC	CAAGTATACA	AACAAAGCCA	TTCCGCCTTA
	301	ACCGCCTTTC	AGACCGAGCA	AATACAAGAT	TCGGAGCATT	CCGGGAAGAT
	351	GGTTGCGAAA	CGCCAGTTCA	GAATCGGCGA	CATAGCGGGC	GAACATACAT
60	401	CTTTTGACAA	GCTTCCCGAA	GGCGGCAGGG	CGACATATCG	CGGGACGGCG
	451	TTCGGTTCAG	ACGATGCCGG	CGGAAAACTG	ACCTACACCA	TAGATTTCGC
	501	CGCCAAGCAG	GGAAACGGCA	AAATCGAACA	TTTGAAATCG	CCAGAACTCA
	551	ATGTCGACCT	GGCCGCCGCC	GATATCAAGC	CGGATGGAAA	ACGCCATGCC
	601	GTCATCAGCG	GTTCCGTCCT	TTACAACCAA	GCCGAGAAAG	GCAGTTACTC
65	651			AAGCCCAGGA		
	701			CGCCATATCG		
	751			ATCCGCCACA		

WO 01/64922

-51-

PCT/IB01/00452

	801	AGCTGCCACT	GTGGCCATTG	CTGCTGCCTA	CAACAATGGC	CAAGAAATCA
	851	ACGGTTTCAA	AGCTGGAGAG	ACCATCTACG	ACATTGATGA	AGACGGCACA
	901	ATTACCAAAA	AAGACGCAAC	TGCAGCCGAT	GTTGAAGCCG	ACGACTTTAA
	951				CCTGACCAAA	
5	1001				AAGCTGCAGA	
	1051				GATGCCGCTT	
	1101				CTTGAATAAA	
	1151				CAAATATCGT	
10	1201				GACAAGCATG	
10	1251				CAACACTAAG	
	1301				CGGCCGAAGA	
	1351				ACTGCAGCAG	
	1401				CGACAAGGCC	
15	1451				TCGCTACGAA	
13	1501				TACACCAGAG	
	1551				CGCTACTACC	
	1601				TTGCCGATCA	
	1651				CTGCGCAAAG	
20	1701				TCTGTTCCAA	CCTTACAACG
20	1751	TGGGTCTCGA	GCACCACCAC	CACCACCACT	GA	
	1	አለር የአ የነገር አርተ	ארט אר תוא דו ארט א	TENTE OF THE	DOCTOVATOV	TET N N OO N TOTETT
	1 51			~	DQSVRKNEKL QLITLESGEF	~
	101				QLITLESGEF EHTSFDKLPE	
25	151				PELNVDLAAA	
23	201		_		SAEVKTVNGI	
	251			~	QEINGFKAGE	,-
	301				TVNENKQNVD	
	351				LGENITTFAE	
30	401				ADEAVKTANE	
	451				EAVAAKVTDI	
	501				EKLDTRLASA	
	551				PYNVGLEHHH	
~ =						
35						
	<u>∆G741—9</u>		~~~~~~~		~~~~~~~~~	m
	1	ATGGTCGCCG	CCGACATCGG		GCCGATGCAC	
	- 4	~~~~~~~			TTTGACGCTG	GATCAGTCCG
	51		AAAGACAAAG		~~~~~~~	
40	101	TCAGGAAAAA	CGAGAAACTG	AAGCTGGCGG	CACAAGGTGC	GGAAAAAACT
40	101 151	TCAGGAAAAA TATGGAAACG	CGAGAAACTG GTGACAGCCT	AAGCTGGCGG CAATACGGGC	AAATTGAAGA	GGAAAAAACT ACGACAAGGT
40	101 151 201	TCAGGAAAAA TATGGAAACG CAGCCGTTTC	CGAGAAACTG GTGACAGCCT GACTTTATCC	AAGCTGGCGG CAATACGGGC GCCAAATCGA	AAATTGAAGA AGTGGACGGG	GGAAAAAACT ACGACAAGGT CAGCTCATTA
40	101 151 201 251	TCAGGAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA	AAATTGAAGA AGTGGACGGG AACAAAGCCA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA
40	101 151 201 251 301	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT
	101 151 201 251 301 351	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT
40	101 151 201 251 301 351 401	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG
	101 151 201 251 301 351 401 451	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC
	101 151 201 251 301 351 401 451 501	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA
	101 151 201 251 301 351 401 451 501	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA GGCCGCCCCC	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC
45	101 151 201 251 301 351 401 451 501 551 601	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA GGCCGCCCCC	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC
	101 151 201 251 301 351 401 451 501 551 601	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA GGCCGCCCC GTTCCGTCCT TTTGGCGGAA	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG
45	101 151 201 251 301 351 401 451 501 551 601 651	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG ATGTCAGCAG ATGTCAGCCT GTCATCAGCCT CCTCGGTATC TGAAAACCGT	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA GGCCGCCGCC GTTCCGTCCT TTTGGCGGAA AAACGGCATA	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC GCCTTGCCGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC
45	101 151 201 251 301 351 401 451 501 551 601 651 701	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG GTCATCAGCC GTCATCAGCG CCTCGGTATC TGAAAACCGT GAGGGATCCG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCGCC GTTCCGTCCT TTTGGCGGAA AAACGGCATA GCGGAGGCGG	AAGCTGGCGG CAATACGGCG GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGC CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC GCCTTGCCGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCG
45	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TCCGGTTCAG CGCCAAGCAG GTCATCAGCC GTCATCAGCG CCTCGGTATC TGAAAACCGT GAGGGATCCG TACCGGTATC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCCC GTTCCGTCCT TTTGGCGGAA AAACGGCATA GCGGAGGCGG GGCAGCACA	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGC CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC GCCTTGCCGC CCCGACTTCA AACAGCGAAA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCAG
45 50	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT GAGGGATCCG TACCGGTATC TATCTTACGC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGA ACGATGCCGC GGAAACGGCA TTTGGCGGAA AAACGGCAT AAACGGCATA GCGGAGCACA CGGCAGCAACA CGGCAGCAACA CGGTATCAAG	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC AACGAAATGT	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCAAGAAAG AGTTGCCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCAGACTC ATGCAGGCGG AAGCAACTC
45	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT GAGGGATCCG TACCGGTATC TACCGGTATC TATCTTACGC TGTGCCGGTC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGA ACGATGCCGC GGCCGCCCGCTTCCGTCCT TTTGGCGAA AAACGGCATA ACGGCATA GCGGAGCAGCA CGGCAGCAACA CGGTATCAAG GGGATGACGT	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC AACGAAATGT TGCGGTTACA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCGGAAG ATGCAGCAGCAG AAGCATGCTC CCAAAATCAA
45 50	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC CTGAGAACCGT TGAGAGATCCG TACCGGTATC TATCTTACGC TATCTTACGC TGTGCCGGTC TGCCCCCCC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCC TTTGGCGGAA AAACGGCATA GCGGAGCAGCA GGCAGCAGCACA CGGCAGCAGCACACA CGGTATCAAG GGCATGACGT CCGAATCTGC	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC AACGAAATGT TGCGGTTACA ATACCGGAGA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG GCCTTGCCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG GCAAAGACAC GACAGGGATG CTTTCCAAAC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCGG AAGCATCC CAGAACTC CCAAAATCAA CCAAATGACG
45 50	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT TGAGAGATCCG TACCGGTATC TGAGGGATCCG TACCGGTATC TATCTTACGC TGTGCCGGTC TGTGCCCCCCC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCC TTTCGCGGAA AAACGGCATA GCGGAGCACA GGCAGCACA CGGCAGCACA CGGTATCAAG GGCAGCACA CGGTATCAAG CGGATGCCT CCGAATCTGC TTTGATCAAC	AAGCTGGCGG CAATACGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCCAAC AACGAAATGT TGCGGTTACA ATACCGGAGA CTCAAACCTG	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG GCCTGCCGC CCCGACTTCCA AACAGCGAAA GCAAAGACAG GAAAGACAG GACAGGGATG CTTTCCAAAC CAATTGAAGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCAG CAAGCAACTC CCAAAATCAA CCAAATGACG AGGCTATACA
45 50	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CTCGGTATC TGAAAACCGT TGAGAATCCG TACCGGTATC TATCTTACGC TACCGGTATC TGTGCCCGCTC CTTGCCCCCCC CATACAAGAA GGACGCGGGG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCC GTTCCGTCCT TTTGGCGGAA ACACGCATA GCGAGCAGC GGCAGCACA GGCAGCACA CGGTATCAAC GGCATGACGT CCGAATCTGC TTTGATCAAC TAGAGGTAGC	AAGCTGGCGG CAATACGGCC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG GCAGAGCAAC GACTTCTGCG GCAGAGCAAC AACGAAATGT TGCGGTTACA ATACCGGAGA CTCAAACCTG TATCGTCGC	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG GCCTTCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG GACAGGGATG CATTCCAAAC CAATTGAAGC ACAGGCGAAT	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CCAGCACTC ATGCAGCAGCG TCAGCAGCAC CCAGAACTC ATGCAGCAGC AGCATGCT CCAAATCAA CCAAATCAA CCAAATGACG AGCCTATACA CCGTCGGCAG
455055	101 151 201 251 301 351 401 451 501 651 701 751 801 851 901 951 1001 1051	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT TACCGGTATC TACCGGTATC TACCGGTATC TATCTTACGC CTGTGCCCCCC CATACAAGAA GGACGCGGGG CATATCCTTT	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA TTTTGGCGGAA AAACGGCATA ACGGCATA ACGGCAGCA GCGAGCAGC GCGAGCACA CGGTATCAAG GCGATGACGT TTTGATCAAC TAGAGGTAGG CCCGAACTGT	AAGCTGGCGG CAATACGGCC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG GCAGAGCAAC TGCGGAGACAC AACGGAAATGT TGCGGTTACA ATACCGGAGA CTCAAACCTG TATCGTCGAC ATGGCAGAAA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC CCCGACTTCA AACAGCGAAA GCAAGGGAAA GCAAGGGAAC CAATTGAAGC CAATTGAAGC ACAGCGAAT CAATTGAAGC ACAGCGAAT AGAACACGGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGCAGG ATGCAGCAG AAGCATGCT CCAAAATCAA CCAAATGACG AGGCTATACA CCGTCGGCAG TATAACGAAA
45 50	101 151 201 251 301 351 401 451 501 551 601 751 801 851 901 951 1001 1051 1101	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT TACCGGTATC TACCGGTATC TATCTTACGC CTGTGCCCCCC CATACAAGAA GGACGCGGGG CATATCCTTT ATTACAAAAA	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGC GGAAACGGCA TTTGGCGGAA AAACGGCATA ACGGAGCAG GGCAGCACA GGCAGCACAC CGGATCCAC CGGATCCAC TTTGACAGC TTTGATCAAC TAGAGGTAGC CCCGAACTGT CTATACGCCG	AAGCTGGCGG CAATACGGCC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA GATATCAAGC TTACAACCAA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC TGCGGAAAATGT TGCGGTTACA ATACCGGAGA CTCAAACCTG TATCGTCGAC ATGGCAGAAA TATATGCGGA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC CCCGACTTCA AACAGCGAAA GCAAGGGAAA GCAAGGGAAC CAATTGAAGC CAATTGAAGC ACAGCGAAT CAATTGAAGC ACAGCGAAT AGAACACGGC ACAGCGAAT AGAACACGGC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGCAG AAGCAACTC ATGCAGCAG AAGCATGCT CCAAAATCAA CCAAATGACG AGGCTATACA CCGTCGGCAG TATAACGAAA TGAAGACGGA
455055	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 951 1001 1051 1101 1151	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG ATGTCGACCT GTCATCAGCG CCTCGGTATC TGAAAACCGT TACCGGTATC TACCGGTATC TATCTACGC TGCCCCCC CATACAAGAA GGACCGGGG CATATCCTTT ATTACAAAAA GGCGGTAAAG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA TTTGGCGGAA AAACGGCATA GCGGAGCATA GCGGAGCAACA CGGTATCAAG GGGATGACGT TTTGATCAAC TTGATCAAC TAGAGGTAGG CCCGAACTGT CCGAACTGT CCGAACTGT CCGAACTGT CTATACGCCG ACATTGAAGC	AAGCTGGCGG CAATACGGCG GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGG GGCGGCAGGG CGGAAAACTG AAATCGAACA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAGCAAC AACGAAATGT TGCGGTTACA ATACCGGAGA ATACCGACA ATACGTCGAC TATCGTCGAC ATGGCAGAAA TATATGCGGA TTCTTTCGAC	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC CCCGACTTCA AACAGCGAAA GCAAAGACAAG GACAGGGATG CTTTCCAAAC CAATTGAAGC ACAGCGAAT ACAGCGAAT ACAGCGAAT ACAGCGAAT CAATTGAAGC ACAGGCGAAT AGAACACGGC ACAGCGCAAT AGAACACGGC AGGAAGCGCC GATGAGGCCC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCAG CCAAATCAA CCAAATGACG AGGCTATACA CCGTCGGCAG TATAACGAAA TGAAGACGA TTATAGAGAC
455055	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 901 901 901 1001 1101 1151 1201	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCT GTCATCAGCG CTCGGTATC TGAAAACCGT TACCGGTATC TATCTTACCC TATCTTACCC CATACAAGAA GGACGCGGGG CATATCCTTT ATTACAAAAA GGCGGTAAAG GGAGGGTAAAG TGAAGCAAAG TGAAGCAAAG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA GCTTCCGAA ACGATGCCGG GGAAACGGCA TTTGGCGGAA AAACGGCATA GCGGAGCATA GCGGAGCACACA CGGTATCAAG GGGATGACGT TTTGATCAAC TTTGATCAAC CCGAACTGT CCGAACTGT CCGAACTGT CTATACGCCG ACATTGAAGC CCGAACTGA	AAGCTGGCGG CAATACGGCG GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGG GGCGGCAGGG CGGAAAACTG AAATCGAACA AAGCCCAGGA CGCCATATCG GCAGAGCAACT CACTCTGCG GCAGAGCAAC AACGAAATGT TGCGGTTACA ATACCGAGA CTCAAACCTG ATACCGCAGA ATACCGCAGA CTCAAACTG TATCGTCGAC ATGGCAGAAA TATATGCCGA TTCTTTCGAC TCCGCCACGT	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGAGAAAG AGTTGCCGGC CCCGACTTCA AACAGCGAAA GCAAAGACAC GACAGGGATGAAA CAAGGGAAT GAAAGACAC GAAAGCAC CAATTGAAGC CAATTGAAGC ACAGGCGAAT AGAACACGGC ACGAAGCCC AGAAGCCCC AAAAGAAATC	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTCCC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGCAG TCAGCAGCAG ACGCATGCT CCAAAATCAA CCAAATGACG AGGCTATACA CCGTCGGCAG TATAACGAAA TGAAGACGA TTATAGAGAC GGACACATCG
455055	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 951 1001 1051 1101 1151 1201 1251	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG GTCATCAGCG CTCATCAGCG TAAAACCGT TACCGGTATC TATCTTACGC TATCTTACGC CATACAGAA GGACGCGGCC CATACAAGAA GGACGCGGGG CATATCCTTT ATTACAAAAA GGCGGTAAAG TGAAGCAAAG ATTTGGTCTC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTCA AGACCGAGCA GCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCGCC GTTCCGTCCT TTTGGCGGAA AAACGGCATA GCGGAGCACA GGCAGCACACA CGGTATCAAG GGGATGACGT TTTGATCAAC TTTGATCAAC TTGAGGTAGC CCCGAACTGT CTATACGCCG ACATTGAAGC CCGACGGATA CCGACGGATA CCGACGGATA CCGACGGATA CCGACGGATA CCGACGGATA CCGACGGATA	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGC CGGAAAACTG AAATCGAACA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAACATC AACGAAAAT TGCGGTTACA ATCGGAGA CTCAAACCTG ATTCGTCGAC ATTCGTCGAC ATTCGTCGAC ATTCGTCGAC ATTCGTCGAC ATTCTTCGAC ATTCTTCGAC TCCGCCACGT GGCGGGCCTT	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGACATTCA ACTTGCCGGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG CTTTCCAAAC CAATTGAAGC ACTTGAAGC CAATTGAAGC CAATTGAAGC CAATTGAAGC CAATTGAAGC ACAGGCGAAT AGAACACGGC ACAGGCGAAT AGAACACGGC AGGAAGCCCC GATGAGGCCCG AAAAGAAATC CCGTGGACGG	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC AGCGCGGAAG CAAGCAACTC ATGCAGCGG TCAGCAGCAG AGCATGCTC CCAAAATCAA CCAAATGCAG AGGCTATACA CCGTCGGCAG TATAACGAAA TGAAGACGA TTATAGAGAC GGACACTCG CAGACCTGCA
45505560	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101 1151 1201 1251 1301	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG GTCATCAGCG CTCATCAGCG TACAGCAG TACCGGTATC TATCTTACGC TGTCCCCCC CATACAAGAA GGACGCGGG CATACCAAGAA GGACGCGGG CATACCATC TATTACCTT TATTACAAAAA GGCGGTAAAG GGCGGTAAAG GGCGGTAAAG TGAAGCAAAG ATTTGGTCTC GGCGGTATTG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA GGCCGCCCC GTTCCGTCCT TTTGGCGGAA AAACGGCATA GCGGAGCACA CGGTATCAAG GGGATGACGT TTTGATCAAC TTGAGTAGG TTTGATCAAC CCGAACTGT CTATACGCCG ACATTGAAGC CCGACGGATA CCATATTATT	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGC CGGAAAACTG AAATCGAACA AAGCCCAGGA AAGCCCAGGA CGCATATCG CACTTCTGCG GCAGAGCAAC TGCGGTTACA ATACCGAGA ATACCGAGA ATACCGAGA CTCAAACCTG TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TTCTTCGAC TCTTTCGAC TCCGCCACGT GGCGGCGCTT GACGCTACAC	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGACATTCA ACTTGCCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG CTTTCCAAAC CAATTGAAGC ACAGGGAAT AGAACAGCGA AGGAAGCGC AGGAAGCGC AGGAAGCGC AAAAGAAATC CCGTGGACGG ATAATGAATA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC ACGCCGGAAG CAAGCAACTC ATGCAGCAGG TCAGCAGCAG ACGCAGACTC ATGCAGCAG AGCATGCT CCAAAATCAA CCAAATGACA CCGTCGGCAG TTATACGAAA TGAAGACGA TTATAGAGAC GGACACATCG CAGACTTCA CGAACTCCA CGACACTCCA CGACCTGCA CGACCTGCA CGACCTGCA CGACCTGCA CGACCTGCA CGACCTGCA CGAATGATGA
455055	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101 1151 1201 1251 1301 1351	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG GTCATCAGCG CCTCGGTATC TGAAAACCGT TGAGGGATCCG TACCGGTATC TATCTTACGC CATACAAGAA GGACGCGGG CATACCAAGAA GGCGGTATC TATCTTACATT ATTACAAAAA GGCGGTAAAG TGAAGCAAAG ATTTGGTCTC GGCGGTATTC GGCGGTATTC GGCGGTATTC AACCAAGAAC	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTCA AGACCGAGCA GCTTCCCGAA ACGATGCCGG GGAAACGGCA TTTGGCGGAA AAACGGCATA ACGGTATCAAG GGCAGCATCAAC CGGTATCAAG GGGATGACGT TTTGATCAAC TAGAGGTAGC CCGAACTGT CTATACGCCG ACATTGAAGC CCGACGGATA CCATATTATT CGCCCGATGC GAAATGATGG	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGC CGGAAAACTG AAATCGAACA AAGCCCAGGA CGCCATATCG CACTTCTGCG GCAGAAATGT TGCGGTTACA ATACCGAGA CTCAAACCTG TATCGTCGAC ATATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TTGCAGACAA TGTTTTCGAC TCCGCCACGT GCCGCGCGTT GACGCTACAC TTGCAGCCAT	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGACATTCA ACAGCGAAA GCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG CTTTCCAAAC CAATTGAAGC CAATTGAAGC CAATTGAAGC CAATTGAAGC ACAGGCGAAT AGAACACGC AGGAAGCGCC AGAAGACACC GATGAGGCCC CGATGGACCC CAATGAAGCCC AAAAGAAATC CCGTGGACGC ATAATGAATA CCGCCAATGCA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC AGCGCGGAAG CAAGCAACTC ATGCAGGCGG TCAGCAGCAG ACGCATGCT CCAAAATCAA CCAAATGACG AGGCTAGCC ACGCCAGAA TCAAACGAAA TGAAGACGAG TTATACGAAA TGAAGACGAG TTATAGAGAC CGACACTCC CAGACTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGACCTCCA CGAATGATCA CGAATGATCA CGAATGATCA CGAATGATCA CGAATGATCA
45505560	101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101 1151 1201 1251 1301	TCAGGAAAAA TATGGAAACG CAGCCGTTTC CCTTGGAGAG ACCGCCTTTC GGTTGCGAAA CTTTTGACAA TTCGGTTCAG CGCCAAGCAG ATGTCGACCG GTCAGTATC GAAAACCGT TGAAAACCGT TACCGGTATC TATCTTACGC TATCTTACGC CATACAAGAA GGACGCGGGC CATACAAAAA GGCGGTAAAG TTACAAAAA GGCGGTAAAG TGAAGCAAGA ATTTGGTCTC GGCGTATTC GGCGGTATTC ATTACAAAAA TGGCGGTAAAG ATTTGGTCTC GGCGTATTC AACCAAGAAC TGGGCGAACG	CGAGAAACTG GTGACAGCCT GACTTTATCC TGGAGAGTTC AGACCGAGCA CGCCAGTTCA ACGATGCCGG GGAAACGGCA TTTGGCGGAA AAACGGCATA AAACGGCATA ACGATGCAA ACGGTATCAG GGCAGCACAC CGGATCAAC CGGATCAAC CGGATCAAC CCGAATCTC TTTGATCAAC TAGAGGTAGC CCGAACTGT CTATACGCCG ACATTGAAGC CCGACGGATA CCATATTATT CGCCCGATGC GAAATGATGC GAAATGATGG GAAATGATGC CGAATTGATCAC CCGACGGATA CCATATTATT CGCCCGATGC GAAATGATGG	AAGCTGGCGG CAATACGGGC GCCAAATCGA CAAGTATACA AATACAAGAT GAATCGGCGA GGCGGCAGGG CGGAAAACTG AAATCGAACA AATCAAGC TTACAACCA AAGCCCAGGA CGCCATATCG GCAGAGCAAC AACGAAATGT TGCGGTTACA ATACGGAGA CTCAAACCTG TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TATCGTCGAC TTCTTTCGAC TCCGCCACGT GCCGCCACGT GCCGCGCCAT CTCCGCACAT TCCGCCACGT GCCGCCACAT CTCCGCCACGT GCCGCGCCACT TTCCGAC TTTCCGAC TTCCGCCACAT TTCCACC TTCCAGCCAT ACCGTCAATA	AAATTGAAGA AGTGGACGGG AACAAAGCCA TCGGAGCATT CATAGCGGGC CGACATATCG ACCTACACCA TTTGAAATCG CGGATGGAAA GCCGACATTCA ACTTGCCGC CCCGACTTCA AACAGCGAAA GCAAAGACAG GACAGGGATG CTTTCCAAAC CAATTGAAGC ACAGGGAAT AGAACAGCGA AGGAAGCGC AGGAAGCGC AGGAAGCGC AAAAGAAATC CCGTGGACGG ATAATGAATA	GGAAAAACT ACGACAAGGT CAGCTCATTA TTCCGCCTTA CCGGGAAGAT GAACATACAT CGGGACGGCG TAGATTTCGC CCAGAACTCA ACGCCATGCC GCAGTTACTC ATGCAGCAGCG TAGCAGCAGC CAAACTCA ACGCATGCT CCAAAATCAA CCAAATGACG AGGCTATACA CCGTCGGCAG TATACGAAA TGAAGACGA TTATAGAGAC CGACACTCC CAAATGACG TATACGAAA TGAAGACGA TTATAGAGAC CGACACTCC CAAATGACG CAGACCTGCA CGACCTGCA CGAACTCG CAGACCTGCA CGAATGATGA TGGGTCAAGC ACAACATCG

-52-

	1551	CCGCCAAGCG	TTGCTCGACT	ATTCCGGCGG	TGATAAAACA	${\tt GACGAGGGTA}$
	1601		GCAACAGAGC			
	1651		TGCTTTTCAT			
5	1701		TATGCCCTAT			
3	1751		AGTCGCAGGC			
	1801 1851		GAGAACCGGG ACTGCCATGT			
	1901		CCGTACAAAC			
	1951		TAACCGGCAC			
10	2001		GACAACCTGC			
	2051	TCGGTGCAGT	CGGCGTGGAC	AGCAAGTTCG	GCTGGGGACT	GCTGGATGCG
	2101	GGTAAGGCCA	TGAACGGACC	${\tt CGCGTCCTTT}$	CCGTTCGGCG	ACTTTACCGC
	2151	CGATACGAAA	${\tt GGTACATCCG}$	${\tt ATATTGCCTA}$	CTCCTTCCGT	AACGACATTT
1.5	2201		CGGCCTGATC			
15	2251		CCTATACGGG			
	2301		AACAACAAAT			
	2351		CGGGGCGCA			
	2401		CAGATACCGA			
20	2451 2501		CTGCAGCTGG GAAAGTGGAC			
20	2551		GCGGCAAGGG			
	2601		CTGAGTGCCG			
	2651		AACCGACGGC			
	2701		GCAGTGAAGG			
25	2751	CAATGCGGCA	CGGACTGCTT	CGGCAGCGGC	ACATTCCGCG	CCCGCCGGTC
	2801	TGAAACACGC	CGTAGAACAG	GGCGGCAGCA	ATCTGGAAAA	CCTGATGGTC
	2851		CCTCCGAATC			
	2901		CGCACAGATA			
30	2951		GGCAGCCGTA			
30	3001		GTCTCGCCGC			
	3051 3101		CAGGGACGCC GGGTCTGCGC			
	3151		AGGGCGGTGT			
	3201		GCCGCGAAAA			
35	3251		ACGCAGCACA			
	3301	GACAGCATTA	GTCTGTTTGC	AGGCATACGG	CACGATGCGG	GCGATATCGG
	3351	CTATCTCAAA	GGCCTGTTCT	CCTACGGACG	CTACAAAAAC	AGCATCAGCC
	3401	GCAGCACCGG	TGCGGACGAA	CATGCGGAAG	GCAGCGTCAA	CGGCACGCTG
40	3451		GCGCACTGGG			
40	3501		GTCGAAGGCG			
	3551		AAAAGGCAGT			
	3601 3651		TGGTCGGACT GTCCTGTTTG			
	3701		CACGGTAACG			
45	3751		GGGCACGCAA			
	3801		GTCGAATTCG			
	3851		TTCCAAACAG			
	3901	GGCTACCGGT	TCCTCGAGCA	CCACCACCAC	CACCACTGA	
50						
50	_1		ADALTAPLDH			
	51		KLKNDKVSRF			
	101 151		SEHSGKMVAK TYTIDFAAKQ			
	201		AEKGSYSLGI			
55	251		PDFNAGGTGI			
	301		DRDAKINAPP			
	351		TGESVGSISF			
	401	GGKDIEASFD	DEAVIETEAK	PTDIRHVKEI	GHIDLVSHII	GGRSVDGRPA
60	451	GGIAPDATLH	IMNTNDETKN	EMMVAAIRNA	WVKLGERGVR	IVNNSFGTTS
60	501		ANSEEQYRQA			
	551		GNDAQAQPNT			
	601		LEYGSNHCGI			
	651		LLQKYPWMSN			
65	701 751		PFGDFTADTK			
03	751 801		IEGGSLVLYG ANETVHIKGS			
	851		LNSTGRRVPF			
	JJ-			_~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	~~~~~~~	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

5	901 951 1001 1051 1101 1151	ELDASESSAT IFNSLAATVY TWEQGGVEGK DSISLFAGIR	SYYVRRGNAA PETVETAAAD ADSTAAHADM MRGSTQTVGI HDAGDIGYLK VPFAATGDLT	RTDMPGIRPY QGRRLKAVSD AAKTGENTTA GLFSYGRYKN	GATFRAAAAV GLDHNGTGLR AATLGMGRST SISRSTGADE	QHANAADGVR VIAQTQQDGG WSENSANAKT HAEGSVNGTL
10	1201 1251 1301		KLSQPLSDKA TRLVAGLGAD HH*			
10	ΔG741-0	n=16 1				
	1		CCGACATCGG	тесевеестт	GCCGATGCAC	TAACCGCACC
	51		AAAGACAAAG			
	101	TCAGGAAAAA	CGAGAAACTG	AAGCTGGCGG	CACAAGGTGC	GGAAAAAACT
15	151	TATGGAAACG	GTGACAGCCT	CAATACGGGC	${\tt AAATTGAAGA}$	ACGACAAGGT
	201		GACTTTATCC			
	251		TGGAGAGTTC			
	301 351		AGACCGAGCA CGCCAGTTCA			
20	401		GCTTCCCGAA			
20	451		ACGATGCCGG			
	501	CGCCAAGCAG	GGAAACGGCA	AAATCGAACA	TTTGAAATCG	CCAGAACTCA
	551	ATGTCGACCT	GGCCGCCGCC	GATATCAAGC	CGGATGGAAA	ACGCCATGCC
0.5	601		GTTCCGTCCT			
25	651		TTTGGCGGAA			
	701 751		AAACGGCATA GAGGCACTGG			
	801		CTCGACCGTC			
	851		CAGGGGGGAA			
30	901	GGAAAAATAC	AAAGCCATCA	GTTGGGCAAC	CTGATGATTC	AACAGGCGGC
	951		AATATCGGCT			
	1001		CCCCTTCGAC			
	1051		CCGTTGACGG			
35	1101 1151		CATCCCGCCG CAAAGGCGCG			
	1201		ATATCCGCCT			
	1251		GACCGTTTCC			
	1301	TAGGCGACGG	ATTCAAACGC	GCCACCCGAT	ACAGCCCCGA	GCTGGACAGA
40	1351		CCGCCGAAGC			
40	1401		GCGGCAGGAG			
	1451 1501		AGGCTCAAAC ACAAGATGGC			
	1551		TATGCCGCAG			
	1601	-	ACAAGGCATA			
45	1651	ATCCCCATCA	AAGGGATTGG	AGCTGTTCGG	GGAAAATACG	GCTTGGGCGG
	1701	CATCACGGCA	CATCCTATCA	AGCGGTCGCA	GATGGGCGCG	ATCGCATTGC
	1751		ATCCGCCGTC			
	1801		CCCCTTACCA			
50	1851		AAAGAAAACA			
50	1901 1951		CAAACTGGCA AAGGGTTTCC			
	2001		CACCACCACC		AAGCACGIGA	
	1	MVAADIGAGL	ADALTAPLDH	KDKGLOSLTI	DOSVRKNEKI.	KLAAOGAEKT
55	51		KLKNDKVSRF			
	101		SEHSGKMVAK			
	151		TYTIDFAAKQ			
	201		AEKGSYSLGI			
60	251		LANDSFIRQV			
60	301		LMIQQAAIKG			
	351 401		YRIHWDGYEH DNRSTGQRLA			
	451		TADIVKNIIG			
	501		DLADMAQLKD			
65	551		GKYGLGGITA			
	601	KYPSPYHSRN	IRSNLEQRYG	KENITSSTVP		
	651	FDGKGFPNFE	KHVKYDTLEH	ннннн*		

Example 16 – C-terminal fusions ('hybrids') with 287/∆G287

5

10

15

According to the invention, hybrids of two proteins A & B may be either NH₂–A–B–COOH or NH₂–B–A–COOH. The effect of this difference was investigated using protein 287 either C-terminal (in '287-His' form) or N-terminal (in Δ G287 form – sequences shown above) to 919, 953 and ORF46.1. A panel of strains was used, including homologous strain 2996. FCA was used as adjuvant:

	287 & 919		287 & 953		287 & ORF46.1	
Strain	∆G287-919	919-287	∆G287-953	953-287	∆G287-46.1	46.1-287
2996	128000	16000	65536	8192	16384	8192
BZ232	256	128	128	<4	<4	<4
1000	2048	<4	<4	<4	<4	<4
MC58	8192	1024	16384	1024	512	128
NGH38	32000	2048	>2048	4096	16384	4096
394/98	4096	32	256	128	128	16
MenA (F6124)	32000	2048	>2048	32	8192	1024
MenC (BZ133)	64000	>8192	>8192	<16	8192	2048

Better bactericidal titres are generally seen with 287 at the N-terminus (in the ΔG form)

When fused to protein 961 [NH₂- Δ G287-961-COOH – sequence shown above], the resulting protein is insoluble and must be denatured and renatured for purification. Following renaturation, around 50% of the protein was found to remain insoluble. The soluble and insoluble proteins were compared, and much better bactericidal titres were obtained with the soluble protein (FCA as adjuvant):

	2996	BZ232	MC58	NGH38	F6124	BZ133
Soluble	65536	128	4096	>2048	>2048	4096
Insoluble	8192	<4	<4	16	n.d.	n.d.

Titres with the insoluble form were, however, improved by using alum adjuvant instead:

Insoluble	32768	128	4096	>2048	>2048	2048	

Example 17 – N-terminal fusions ('hybrids') to 287

Expression of protein 287 as full-length with a C-terminal His-tag, or without its leader peptide but with a C-terminal His-tag, gives fairly low expression levels. Better expression is achieved using a N-terminal GST-fusion.

As an alternative to using GST as an N-terminal fusion partner, 287 was placed at the C-terminus of protein 919 ('919-287'), of protein 953 ('953-287'), and of proteins ORF46.1 ('ORF46.1-287'). In both cases, the leader peptides were deleted, and the hybrids were direct in-frame fusions.

To generate the 953-287 hybrid, the leader peptides of the two proteins were omitted by designing the forward primer downstream from the leader of each sequence; the stop codon sequence was omitted in the 953 reverse primer but included in the 287 reverse primer. For the 953 gene, the 5' and the 3' primers used for amplification included a *NdeI* and a *BamHI* restriction sites respectively, whereas for the amplification of the 287 gene the 5' and the 3' primers included a *BamHI* and a *XhoI* restriction sites respectively. In this way a sequential directional cloning of the two genes in pET21b+, using *NdeI-BamHI* (to clone the first gene) and subsequently *BamHI-XhoI* (to clone the second gene) could be achieved.

The 919-287 hybrid was obtained by cloning the sequence coding for the mature portion of 287 into the *Xho*I site at the 3'-end of the 919-His clone in pET21b+. The primers used for amplification of the 287 gene were designed for introducing a *Sal*I restriction site at the 5'-and a *Xho*I site at the 3'- of the PCR fragment. Since the cohesive ends produced by the *Sal*I and *Xho*I restriction enzymes are compatible, the 287 PCR product digested with *Sal*I-*Xho*I could be inserted in the pET21b-919 clone cleaved with *Xho*I.

The ORF46.1-287 hybrid was obtained similarly.

15

The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens:

	Mixture with 287	Hybrid with 287
919	32000	16000
953	8192	8192
ORF46.1	128	8192

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained for 919-287 and 953-287:

-56-

	919		953		ORF46.1	
Strain	Mixture	Hybrid	Mixture	Hybrid	Mixture	Hybrid
MC58	512	1024	512	1024	-	1024
NGH38	1024	2048	2048	4096	-	4096
BZ232	512	128	1024	16	-	-
MenA (F6124)	512	2048	2048	32	_	1024
MenC (C11)	>2048	n.d.	>2048	n.d.	-	n.d.
MenC (BZ133)	>4096	>8192	>4096	<16	-	2048

Hybrids of ORF46.1 and 919 were also constructed. Best results (four-fold higher titre) were achieved with 919 at the N-terminus.

Hybrids 919-519His, ORF97-225His and 225-ORF97His were also tested. These gave moderate ELISA fitres and bactericidal antibody responses.

5 Example 18 – the leader peptide from ORF4

As shown above, the leader peptide of ORF4 can be fused to the mature sequence of other proteins (e.g. proteins 287 and 919). It is able to direct lipidation in *E.coli*.

Example 19 - domains in 564

The protein '564' is very large (2073aa), and it is difficult to clone and express it in complete form. To facilitate expression, the protein has been divided into four domains, as shown in figure 8 (according to the MC58 sequence):

Domain	A	В	C	D
Amino Acids	79-360	361-731	732-2044	2045-2073

These domains show the following homologies:

• Domain A shows homology to other bacterial toxins:

- Domain B shows no homology, and is specific to 564.
 - Domain C shows homology to:

-57-

• Domain D shows homology to other bacterial toxins:

```
gb|AAF84995.1|AE004032_14 HA-like secreted protein [Xylella fastidiosa] (29%)
```

Using the MC58 strain sequence, good intracellular expression of 564ab was obtained in the form of GST-fusions (no purification) and his-tagged protein; this domain-pair was also expressed as a lipoprotein, which showed moderate expression in the outer membrane/supernatant fraction.

The b domain showed moderate intracellular expression when expressed as a his-tagged product (no purification), and good expression as a GST-fusion.

The c domain showed good intracellular expression as a GST-fusion, but was insoluble. The d domain showed moderate intracellular expression as a his-tagged product (no purification). The cd protein domain-pair showed moderate intracellular expression (no purification) as a GST-fusion.

Good bactericidal assay titres were observed using the c domain and the bc pair.

15 Example 20 – the 919 leader peptide

5

20

35

The 20mer leader peptide from 919 is discussed in example 1 above:

```
MKKYLFRAAL YGIAAAILAA
```

As shown in example 1, deletion of this leader improves heterologous expression, as does substitution with the ORF4 leader peptide. The influence of the 919 leader on expression was investigated by fusing the coding sequence to the *PhoC* reporter gene from *Morganella morganii* [Thaller *et al.* (1994) *Microbiology* 140:1341-1350]. The construct was cloned in the pET21-b plasmid between the *NdeI* and *XhoI* sites (Figure 9):

```
25 1 MKKYLFRAAL YGIAAAILAA AIPAGNDATT KPDLYYLKNE QAIDSLKLLP
PPPEVGSIQF LNDQAMYEKG RMLRNTERGK QAQADADLAA GGVATAFSGA
101 FGYPITEKDS PELYKLLTNM IEDAGDLATR SAKEHYMRIR PFAFYGTETC
151 NTKDQKKLST NGSYPSGHTS IGWATALVLA EVNPANQDAI LERGYQLGQS
201 RVICGYHWQS DVDAARIVGS AAVATLHSDP AFQAQLAKAK QEFAQKSQK*
```

The level of expression of PhoC from this plasmid is >200-fold lower than that found for the same construct but containing the native PhoC signal peptide. The same result was obtained even after substitution of the T7 promoter with the *E.coli* Plac promoter. This means that the influence of the 919 leader sequence on expression does not depend on the promoter used.

In order to investigate if the results observed were due to some peculiarity of the 919 signal peptide nucleotide sequence (secondary structure formation, sensitivity to RNAases, etc.) or

10

45

to protein instability induced by the presence of this signal peptide, a number of mutants were generated. The approach used was a substitution of nucleotides of the 919 signal peptide sequence by cloning synthetic linkers containing degenerate codons. In this way, mutants were obtained with nucleotide and/or amino acid substitutions.

5 Two different linkers were used, designed to produce mutations in two different regions of the 919 signal peptide sequence, in the first 19 base pairs (L1) and between bases 20-36 (S1).

The alignment of some of the mutants obtained is given below.

L1 mutants: 9L1-a 15 9L1-е 9L1-d 9L1-f ATGAAAAATATCTCTTTAGCGCCGCCCTGTACGGCATCGCCGCCGCCATCCTCGCCGCC 919sp ATGAAAAATACCTATTCCGCGCCGCCCTGTACGGCATCGCCGCCCATCCTCGCCGCC 20 9L1a MKKYLFSAA~~~~~~~ 9L1e MKKYFFRAA~~~~~~~ 9L1d MKKYFFRAA~~~~~~~ 9L1f MKKYLFSAALYGIAAAILAA 919sp MKKYLFRAALYGIAAAILAA (i.e. native signal peptide) 25 S1 mutants: ATGAAAAATACCTATTC.....ATCGCCGCCGCCATCCTCGCCGCC 9S1-c ATGAAAAATACCTATTCCGAGCTGCCCAATACGGCATCGCCGCCGCCATCCTCGCCGCC 9s1-b ATGAAAAATACCTATTCCGGGCCGCCCAATACGGCATCGCCGCCCCATCCTCGCCGCC 30 9S1-i ATGAAAAATACCTATTCCGGGCGGCTTTGTACGGGATCGCCGCCCCATCCTCGCCGCC 919sp ATGAAAAATACCTATTCCGCGCCGCCTGTACGGCATCGCCGCCCATCCTCGCCGCC 9S1e MKKYLF.....IAAAILAA 9S1c MKKYLFRAAQYGIAAAILAA 35 9s1b MKKYLFRAAQYGIAAAILAA 9sli MKKYLFRAALYGIAAAILAA 919sp MKKYLFRAALYGIAAAILAA

As shown in the sequences alignments, most of the mutants analysed contain in-frame deletions which were unexpectedly produced by the host cells.

Selection of the mutants was performed by transforming *E. coli* BL21(DE3) cells with DNA prepared from a mixture of L1 and S1 mutated clones. Single transformants were screened for high PhoC activity by streaking them onto LB plates containing 100 μg/ml ampicillin, 50μg/ml methyl green, 1 mg/ml PDP (phenolphthaleindiphosphate). On this medium PhoC-producing cells become green (Figure 10).

A quantitative analysis of PhoC produced by these mutants was carried out in liquid medium using pNPP as a substrate for PhoC activity. The specific activities measured in cell extracts and supernatants of mutants grown in liquid medium for 0, 30, 90, 180 min. were:

CELL EXTRACTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	1,11	1,11	3,33	4,44
9S1e	102,12	111,00	149,85	172,05
9L1a	206,46	111,00	94,35	83,25
9L1d	5,11	4,77	4,00	3,11
9L1f	27,75	94,35	82,14	36,63
9S1b	156,51	111,00	72,15	28,86
9S1c	72,15	33,30	21,09	14,43
9S1i	156,51	83,25	55,50	26,64
phoCwt	194,25	180,93	149,85	142,08

SUPERNATANTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	0,33	0,00	0,00	0,00
9S1e	0,11	0,22	0,44	0,89
9L1a	4,88	5,99	5,99	7,22
9L1d	0,11	0,11	0,11	0,11
9L1f	0,11	0,22	0,11	0,11
9S1b	1,44	1,44	1,44	1,67
9S1c	0,44	0,78	0,56	0,67
9S1i	0,22	0,44	0,22	0,78
phoCwt	34,41	43,29	87,69	177,60

Some of the mutants produce high amounts of PhoC and in particular, mutant 9L1a can secrete PhoC in the culture medium. This is noteworthy since the signal peptide sequence of this mutant is only 9 amino acids long. This is the shortest signal peptide described to date.

Example 21 – C-terminal deletions of Maf-related proteins

MafB-related proteins include 730, ORF46 and ORF29.

The 730 protein from MC58 has the following sequence:

	1.	VKPLRRLTNL	LAACAVAAAA	LIQPALAADL	AQDPFITDNA	QRQHYEPGGK
15	51	YHLFGDPRGS	VSDRTGKINV	IQDYTHQMGN	LLIQQANING	TIGYHTRFSG
	101	HGHEEHAPFD	NHAADSASEE	KGNVDEGFTV	YRLNWEGHEH	HPADAYDGPK
	151	GGNYPKPTGA	RDEYTYHVNG	TARSIKLNPT	DTRSIRQRIS	DNYSNLGSNF
	201	SDRADEANRK	MFEHNAKLDR	WGNSMEFING	VAAGALNPFI	SAGEALGIGD
••	251	ILYGTRYAID	KAAMRNIAPL	PAEGKFAVIG	GLGSVAGFEK	NTREAVDRWI
20	301	QENPNAAETV	EAVFNVAAAA	KVAKLAKAAK	PGKAAVSGDF	ADSYKKKLAL

5

10

-60-

PCT/IB01/00452

- 351 SDSARQLYQN AKYREALDIH YEDLIRRKTD GSSKFINGRE IDAVTNDALI
- 401 QAKRTISAID KPKNFLNQKN RKQIKATIEA ANQQGKRAEF WFKYGVHSQV
- 451 KSYIESKGGI VKTGLGD*
- 5 The leader peptide is underlined.

WO 01/64922

15

30

730 shows similar features to ORF46 (see example 8 above):

- as for Orf46, the conservation of the 730 sequence among MenB, MenA and gonococcus is high (>80%) only for the N-terminal portion. The C-terminus, from ~340, is highly divergent.
- 10 its predicted secondary structure contains a hydrophobic segment spanning the central region of the molecule (aa. 227-247).
 - expression of the full-length gene in E. coli gives very low yields of protein. Expression from tagged or untagged constructs where the signal peptide sequence has been omitted has a toxic effect on the host cells. In other words, the presence of the full-length mature protein in the cytoplasm is highly toxic for the host cell while its translocation to the periplasm (mediated by the signal peptide) has no detectable effect on cell viability. This "intracellular toxicity" of 730 is particularly high since clones for expression of the leaderless 730 can only be obtained at very low frequency using a recA genetic background (E. coli strains: HB101 for cloning; HMS174(DE3) for expression).
- To overcome this toxicity, a similar approach was used for 730 as described in example 8 for ORF46. Four C-terminal truncated forms were obtained, each of which is well expressed. All were obtained from intracellular expression of His-tagged leaderless 730.
 - Form A consists of the N-terminal hydrophilic region of the mature protein (aa. 28-226). This was purified as a soluble His-tagged product, having a higher-than-expected MW.
- Form B extends to the end of the region conserved between serogroups (aa. 28-340). This was purified as an insoluble His-tagged product.

The C-terminal truncated forms named C1 and C2 were obtained after screening for clones expressing high levels of 730-His clones in strain HMS174(DE3). Briefly, the pET21b plasmid containing the His-tagged sequence coding for the full-length mature 730 protein was used to transform the *recA* strain HMS174(DE3). Transformants were obtained at low frequency which showed two phenotypes: large colonies and very small colonies. Several large and small colonies were analysed for expression of the 730-His clone. Only cells from large colonies over-expressed a protein recognised by anti-730A antibodies. However the

5

30

35

protein over-expressed in different clones showed differences in molecular mass. Sequencing of two of the clones revealed that in both cases integration of an *E. coli* IS sequence had occurred within the sequence coding for the C terminal region of 730. The two integration events have produced in-frame fusion with 1 additional codon in the case of C1, and 12 additional codons in the case of C2 (Figure 11). The resulting "mutant" forms of 730 have the following sequences:

```
730-C1 (due to an IS1 insertion - figure 11A)

1 MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
51 QMGNLLIQQA NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
10 101 GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
151 LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANKKMFEHNA KLDRWGNSME
201 FINGVAAGAL NPFISAGEAL GIGDLLYGTR YAIDKAAMRN IAPLPAEGKF
251 AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
301 KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LDIHYEDLIR
15 351 RKTDGSSKFI NGREIDAVTN DALIQAR*
```

The additional amino acid produced by the insertion is underlined.

```
730-C2 (due to an IS5 insertion - Figure 11B)

1 MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
20 51 QMGNLLIQQA NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
101 GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
151 LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANRKMFEHNA KLDRWGNSME
201 FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIDKAAMRN IAPLPAEGKF
251 AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
25 301 KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LGKVRISGEI
```

The additional amino acids produced by the insertion are underlined.

In conclusion, intracellular expression of the 730-C1 form gives very high level of protein and has no toxic effect on the host cells, whereas the presence of the native C-terminus is toxic. These data suggest that the "intracellular toxicity" of 730 is associated with the C-terminal 65 amino acids of the protein.

Equivalent truncation of ORF29 to the first 231 or 368 amino acids has been performed, using expression with or without the leader peptide (amino acids 1-26; deletion gives cytoplasmic expression) and with or without a His-tag.

Example 22 - domains in 961

As described in example 9 above, the GST-fusion of 961 was the best-expressed in *E.coli*. To improve expression, the protein was divided into domains (figure 12).

The domains of 961 were designed on the basis of YadA (an adhesin produced by *Yersinia* which has been demonstrated to be an adhesin localized on the bacterial surface that forms

oligomers that generate surface projection [Hoiczyk et al. (2000) EMBO J 19:5989-99]) and are: leader peptide, head domain, coiled-coil region (stalk), and membrane anchor domain.

These domains were expressed with or without the leader peptide, and optionally fused either to C-terminal His-tag or to N-terminal GST. *E.coli* clones expressing different domains of 961 were analyzed by SDS-PAGE and western blot for the production and localization of the expressed protein, from over-night (o/n) culture or after 3 hours induction with IPTG. The results were:

	Total lysate Periplasm Supernat		Supernatant	OMV
	(Western Blot)	(Western Blot)	(Western Blot)	SDS-PAGE
961 (o/n)	-	-		
961 (IPTG)	+/-	-	-	
961-L (o/n)	+	-	-	+
961-L (IPTG)	+	-	-	+
961c-L (o/n)	-	-	-	
961c-L (IPTG)	+	+	+	
961Δ ₁ -L (o/n)	-	-	-	
$961\Delta_1$ -L (IPTG)	+	-	_	+

The results show that in *E.coli*:

5

- 961-L is highly expressed and localized on the outer membrane. By western blot analysis two specific bands have been detected: one at ~45kDa (the predicted molecular weight) and one at ~180kDa, indicating that 961-L can form oligomers. Additionally, these aggregates are more expressed in the over-night culture (without IPTG induction). OMV preparations of this clone were used to immunize mice and serum was obtained. Using overnight culture (predominantly by oligomeric form) the serum was bactericidal; the IPTG-induced culture (predominantly monomeric) was not bactericidal.
 - 961 Δ_1 -L (with a partial deletion in the anchor region) is highly expressed and localized on the outer membrane, but does not form oligomers;
 - the 961c-L (without the anchor region) is produced in soluble form and exported in the supernatant.
- 20 Titres in ELISA and in the serum bactericidal assay using His-fusions were as follows:

	ELISA	Bactericidal
961a (aa 24-268)	24397	4096

961b (aa 269-405)	7763	64
961c-L	29770	8192
961c (2996)	30774	>65536
961c (MC58)	33437	16384
961d	26069	>65536

E.coli clones expressing different forms of 961 (961, 961-L, 961 Δ_1 -L and 961c-L) were used to investigate if the 961 is an adhesin (c.f. YadA). An adhesion assay was performed using (a) the human epithelial cells and (b) E.coli clones after either over-night culture or three hours IPTG induction. 961-L grown over-night (961 Δ_1 -L) and IPTG-induced 961c-L (the clones expressing protein on surface) adhere to human epithelial cells.

961c was also used in hybrid proteins (see above). As 961 and its domain variants direct efficient expression, they are ideally suited as the N-terminal portion of a hybrid protein.

Example 23 – further hybrids

5

10

Further hybrid proteins of the invention are shown below (see also Figure 14). These are advantageous when compared to the individual proteins:

	ORF46.1	741				
	1	ATGTCAGATT	TGGCAAACGA	TTCTTTTATC	CGGCAGGTTC	TCGACCGTCA
	51	GCATTTCGAA	CCCGACGGGA	AATACCACCT	ATTCGGCAGC	AGGGGGGAAC
	101	TTGCCGAGCG	CAGCGGCCAT	ATCGGATTGG	GAAAAATACA	AAGCCATCAG
15	151	TTGGGCAACC	TGATGATTCA	ACAGGCGGCC	ATTAAAGGAA	ATATCGGCTA
	201	CATTGTCCGC	TTTTCCGATC	ACGGGCACGA	AGTCCATTCC	CCCTTCGACA
	251	ACCATGCCTC	ACATTCCGAT	TCTGATGAAG	CCGGTAGTCC	CGTTGACGGA
	301	TTTAGCCTTT	ACCGCATCCA	TTGGGACGGA	TACGAACACC	ATCCCGCCGA
	351	CGGCTATGAC	GGGCCACAGG	GCGGCGGCTA	TCCCGCTCCC	AAAGGCGCGA
20	401	GGGATATATA	CAGCTACGAC	ATAAAAGGCG	TTGCCCAAAA	TATCCGCCTC
	451				CGGCTTGCCG	
	501	CAATGCCGGT	AGTATGCTGA	CGCAAGGAGT	AGGCGACGGA	
	551	CCACCCGATA	CAGCCCCGAG	CTGGACAGAT	CGGGCAATGC	CGCCGAAGCC
	601				ATCATCGGCG	
25	651				CATAAGCGAA	
	701				CCACCGAAAA	
	751	CGCATCAACG	ATTTGGCAGA		CTCAAAGACT	
	801	AGCCATCCGC	GATTGGGCAG		CAATGCCGCA	
	851	AAGCCGTCAG	CAATATCTTT		TCCCCATCAA	
30	901	GCTGTTCGGG	GAAAATACGG		ATCACGGCAC	
	951	GCGGTCGCAG	ATGGGCGCGA	-	GAAAGGGAAA	
	1001	GCGACAATTT			AATACCCGTC	
	1051	TCCCGAAATA			CGTTACGGCA	
	1101				CAAAAATGTC	
35	1151				TTGACGGTAA	
	1201				GGATCCGGAG	
	1251	CGCCGCCGAC	ATCGGTGCGG		TGCACTAACC	
	1301		CAAAGGTTTG		CGCTGGATCA	
	1351				GGTGCGGAAA	
40	1401	AAACGGTGAC	AGCCTCAATA	CGGGCAAATT	GAAGAACGAC	AAGGTCAGCC
	1451				ACGGGCAGCT	
	1501				AGCCATTCCG	
	1551				GCATTCCGGG	
	1601	CGAAACGCCA	GTTCAGAATC	GGCGACATAG	CGGGCGAACA	TACATCTTTT

-64-

WO 01/64922

PCT/IB01/00452

	1651	CACAACCTTC	CCGAAGGCGG	CAGGGCGACA	ТАТССССССА	СССССТТССС
	1701			AACTGACCTA		
	1751			GAACATTTGA		
	1801	GACCTGGCCG	CCGCCGATAT	CAAGCCGGAT	GGAAAACGCC	ATGCCGTCAT
5	1851	CAGCGGTTCC	${\tt GTCCTTTACA}$	ACCAAGCCGA	GAAAGGCAGT	TACTCCCTCG
	1901			${\tt CAGGAAGTTG}$		
	1951			TATCGGCCTT	GCCGCCAAGC	AACTCGAGCA
	2001	CCACCACCAC	CACCACTGA			
10	1	MODE ANDOUT	DOM DROHEE	PDGKYHLFGS	DOET VEDCOU	TOLOVIOCHO
10	51			FSDHGHEVHS		
	101			GPOGGGYPAP		
	151			SMLTQGVGDG		_
	201	~		AGDAVQGISE		
15	251	RINDLADMAQ	LKDYAAAAIR	DWAVQNPNAA	QGIEAVSNIF	MAAIPIKGIG
	301	AVRGKYGLGG	ITAHPIKRSQ	${\tt MGAIALPKGK}$	SAVSDNFADA	AYAKYPSPYH
	351			TVPPSNGKNV		
	401			IGAGLADALT		
20	451 501			SLNTGKLKND		
20	501 551			EQIQDSEHSG AGGKLTYTID		
	601			VLYNQAEKGS	-	
	651		AAKQLEHHHH		1011011 00141	QLIVI3CDI3LIVIC
25						
	ORF46.1-					
	1			TTCTTTTATC		
	51			AATACCACCT		
30	101 151			ATCGGATTGG		
30	201			ACAGGCGGCC ACGGGCACGA		
	251			TCTGATGAAG		
	301			TTGGGACGGA		
	351			GCGGCGGCTA		
35	401	GGGATATATA	CAGCTACGAC	ATAAAAGGCG	TTGCCCAAAA	TATCCGCCTC
	451			CACCGGACAA		
	501			CGCAAGGAGT		
	551			CTGGACAGAT		
40	601 651			CGTTAAAAAC		
40	701			GGTCTGCTTT		
	751			TATGGCGCAA		
	801			TCCAAAACCC		
	851			ATGGCAGCCA		
45	901			CTTGGGCGGC		
	951			TCGCATTGCC		
	1001			GCATACGCCA		
	1051 1101			CTTGGAGCAG		
50	1151			CGTCAAACGG GGCGTACCGT		
	1201			ATATGATACG		
	1251			AAAAAGCTGC		
	1301			ATCAACGGTT		
	1351	TACGACATTG	ATGAAGACGG	CACAATTACC	AAAAAAGACG	CAACTGCAGC
55	1401			TTAAAGGTCT		
	1451			AATGAAAACA		
	1501			AATAGAAAAG		
	1551			ATACTGATGC		
60	1601 1651			GAAAATATAA TGATGAAAAA		
	1701			CATTCAACGA		
	1751			GAAGCCGTCA		
	1801			ACAAAACGTC		
<i>-</i> =	1851			CCGAAGCTGC		
65	1901			GTCGCTGCAA		
	1951			TAATATTGCT		
	2001	CGTGTACACC	AGAGAAGAGT	CTGACAGCAA	ATTTGTCAGA	ATTGATGGTC

	2051	ТGA ACGCТAC	TACCGAAAAA	TTGGACACAC	GCTTGGCTTC	TGCTGAAAA
	2101				GGTTTGGATA	
	2151				TGCAGAACAA	
	2201				GGTTCAATGT	
5	2251				GCCATCGGTA	
	2301				CGTGGCAGTC	
	2351	CCGGTTCTTC	CGCAGCCTAC	CATGTCGGCG	TCAATTACGA	GTGGCTCGAG
	2401	CACCACCACC	ACCACCACTG	A		
1.0						
10	_1				RGELAERSGH	
	51.	~~			PFDNHASHSD	
	101				KGARDIYSYD	
	151				FKRATRYSPE	
15	201			_	GSNIAVMHGL	
15	251	~		_	QGIEAVSNIF	
	301				SAVSDNFADA	
	351				KLADQRHPKT IAAAYNNGQE	
	401 451				KVVTNLTKTV	
20	501				ATTNALNKLG	
20	551				SLDETNTKAD	
	601				ANTAADKAEA	
	651				IDGLNATTEK	
	701				AALSGLFQPY	
25	751				GTSSGSSAAY	
23	801	HHHHHH*	AIGIGEREIL	M. WHIWGAYA	GIDDGBDAAI	HAGANTEMPE
	302					
		•				
20	ORF46.1					
30	1				CGGCAGGTTC	
	51				ATTCGGCAGC	
	101				GAAAAATACA	
	151				ATTAAAGGAA	
35	201				AGTCCATTCC	
33	251				CCGGTAGTCC	
	301				TACGAACACC TCCCGCTCCC	
	351 401				TTGCCCAAAA	
	451				CGGCTTGCCG	
40	501				AGGCGACGGA	
10	551				CGGGCAATGC	
	601				ATCATCGGCG	
	651				CATAAGCGAA	
	701				CCACCGAAAA	
45	751				CTCAAAGACT	
	801				CAATGCCGCA	
	851				TCCCCATCAA	
	901	GCTGTTCGGG	GAAAATACGG	CTTGGGCGGC	ATCACGGCAC	ATCCTATCAA
	951				GAAAGGGAAA	
50	1001				AATACCCGTC	
	1051				CGTTACGGCA	
	1101	CACCTCCTCA	ACCGTGCCGC	CGTCAAACGG	CAAAAATGTC	AAACTGGCAG
	1151	ACCAACGCCA	CCCGAAGACA	GGCGTACCGT	TTGACGGTAA	AGGGTTTCCG
	1201	AATTTTGAGA	AGCACGTGAA	ATATGATACG	GGATCCGGAG	GAGGAGGAGC
55	1251	CACAAACGAC	GACGATGTTA	AAAAAGCTGC	CACTGTGGCC	ATTGCTGCTG
	1301	CCTACAACAA	TGGCCAAGAA	ATCAACGGTT	TCAAAGCTGG	AGAGACCATC
	1351	TACGACATTG	ATGAAGACGG	CACAATTACC	AAAAAAGACG	CAACTGCAGC
	1401	CGATGTTGAA	GCCGACGACT	TTAAAGGTCT	GGGTCTGAAA	AAAGTCGTGA
	1451	CTAACCTGAC	CAAAACCGTC	AATGAAAACA	AACAAAACGT	CGATGCCAAA
60	1501				TTAACAACCA	
	1551	CACTGATGCC	GCTTTAGCAG	ATACTGATGC	CGCTCTGGAT	GCAACCACCA
	1601	ACGCCTTGAA	TAAATTGGGA	GAAAATATAA	CGACATTTGC	TGAAGAGACT
	1651				TTAGAAGCCG	
~=	1701				TATCGCCGAT	
65	1751				AAACCGCCAA	
	1801				GATGCCAAAG	
	1851	AGAAACTGCA	GCAGGCAAAG	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG

-66-

	1901	CAGCCGACAA	GGCCGAAGCT	GTCGCTGCAA	AAGTTACCGA	CATCAAAGCT
	1951	GATATCGCTA	CGAACAAAGA	TAATATTGCT	AAAAAAGCAA	ACAGTGCCGA
	2001	CGTGTACACC	AGAGAAGAGT	CTGACAGCAA	ATTTGTCAGA	ATTGATGGTC
_	2051			TTGGACACAC		
5	2101			TCGCCTGAAC		
	2151			GCCAAGGCCT		
	2201		CCAACCTTAC	AACGTGGGTC	TCGAGCACCA	CCACCACCAC
	2251	CACTGA				
10	1	MCDI ANDCET	BOXX DDOXEE	DDGKWIII EGG	DOEL VEDCOIL	TOT OWTOOMS
10	51			PDGKYHLFGS FSDHGHEVHS		
	101	~~		GPQGGGYPAP		
	151			SMLTOGVGDG		~
	201	~		AGDAVQGISE		
15	251			DWAVQNPNAA		
	301			MGAIALPKGK		
	351			TVPPSNGKNV		
	401	NFEKHVKYDT	GSGGGGATND	DDVKKAATVA	IAAAYNNGQE	INGFKAGETI
	451	YDIDEDGTIT	KKDATAADVE	ADDFKGLGLK	KVVTNLTKTV	NENKQNVDAK
20	501	VKAAESEIEK	LTTKLADTDA	ALADTDAALD	ATTNALNKLG	ENITTFAEET
	551	KTNIVKIDEK	LEAVADTVDK	HAEAFNDIAD	SLDETNTKAD	EAVKTANEAK
	601	QTAEETKQNV	DAKVKAAETA	AGKAEAAAGT	ANTAADKAEA	VAAKVTDIKA
	651			REESDSKFVR		
05	701		GLDKTVSDLR	KETRQGLAEQ	AALSGLFQPY	NVGLEHHHHH
25	751	H*				
	961-ORE	F46.1				
	1	·······	ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
30	51			AAGAAATCAA		
	101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACCAAAAA	AGACGCAACT
	151	GCAGCCGATG	TTGAAGCCGA	CGACTTTAAA	GGTCTGGGTC	TGAAAAAAGT
	201	CGTGACTAAC	CTGACCAAAA	CCGTCAATGA	AAACAAACAA	AACGTCGATG
25	251			TCTGAAATAG		
35	301			AGCAGATACT		
	351			TGGGAGAAAA		
	401			AAAATTGATG		
	451			CGAAGCATTC		
40	501 551			CAGACGAAGC ACCAAACAAA		
70	601			CAAAGCCGAA		
	651			AAGCTGTCGC		
	701			AAAGATAATA		
	751			AGAGTCTGAC		
45	801			AAAAATTGGA		
	851			GATACTCGCC		
	901	GTGTCAGACC	TGCGCAAAGA	AACCCGCCAA	GGCCTTGCAG	AACAAGCCGC
	951	GCTCTCCGGT	CTGTTCCAAC	CTTACAACGT	GGGTCGGTTC	AATGTAACGG
~ 0	1001	CTGCAGTCGG	CGGCTACAAA	TCCGAATCGG	CAGTCGCCAT	CGGTACCGGC
50	1051			TGCCGCCAAA		
	1101			CCTACCATGT		
	1151			GATTTGGCAA		
	1201			CGAACCCGAC		
55	1251			AGCGCAGCGG		
33	1301 1351			AACCTGATGA CCGCTTTTCC		
	1401			CCTCACATTC		
	1451			CTTTACCGCA		
	1501			TGACGGGCCA		
60	1551			TATACAGCTA		
	1601			ACCGACAACC		
	1651			CGGTAGTATG		
	1701			GATACAGCCC		
	1751			GGCACTGCAG		
65	1801			CGGCGCAGGC		
	1851	CGAAGGCTCA	${\tt AACATTGCTG}$	${\tt TCATGCACGG}$	${\tt CTTGGGTCTG}$	CTTTCCACCG
	1901	AAAACAAGAT	$\tt GGCGCGCATC$	AACGATTTGG	${\tt CAGATATGGC}$	GCAACTCAAA

	1951	GACTATGCCG	CAGCAGCCAT	CCGCGATTGG	GCAGTCCAAA	ACCCCAATGC
	2001	CGCACAAGGC	ATAGAAGCCG	TCAGCAATAT	CTTTATGGCA	GCCATCCCCA
	2051	TCAAAGGGAT	${\tt TGGAGCTGTT}$	CGGGGAAAAT	ACGGCTTGGG	CGGCATCACG
~	2101			GCAGATGGGC		
5	2151			ATTTTGCCGA		
	2201			AATATCCGTT		
	2251	· · ·		CTCAACCGTG		
	2301			GCCACCCGAA		
10	2351			GAGAAGCACG	TGAAATATGA	TACGCTCGAG
10	2401	CACCACCACC	ACCACCACTG	A		
	1	אא זירורורוואייז או	አአመነንአ ተአአአኒና	NNGQEINGFK	ACEMIVATAE	DOWLWRRDAM
	51			LTKTVNENKO		
	101			LNKLGENITT		
15	151			NTKADEAVKT		
10	201			DKAEAVAAKV		
	251			ATTEKLDTRL		
	301			LFQPYNVGRF		
	351			SSAAYHVGVN		
20	401			ELAERSGHIG		
	451	GNIGYIVRFS	DHGHEVHSPF	DNHASHSDSD	EAGSPVDGFS	LYRIHWDGYE
	501	HHPADGYDGP	QGGGYPAPKG	ARDIYSYDIK	GVAQNIRLNL	TDNRSTGQRL
	551	ADRFHNAGSM	LTQGVGDGFK	RATRYSPELD	RSGNÄAEAFN	GTADIVKNII
	601	GAAGEIVGAG	DAVQGISEGS	${\tt NIAVMHGLGL}$	LSTENKMARI	NDLADMAQLK
25	651	DYAAAAIRDW	AVQNPNAAQG	IEAVSNIFMA	AIPIKGIGAV	RGKYGLGGIT
	701			VSDNFADAAY		
	751	GKENITSSTV	PPSNGKNVKL	ADQRHPKTGV	PFDGKGFPNF	EKHVKYDTLE
	801	ннинин*				
30	0.53 5.44					
30	961-741 1	7.maaaaa aa 7			~~T~~~~~~	
	_			TGTTAAAAAA		
	51 101			AAGAAATCAA GACGGCACAA		
	151			CGACTTTAAA		
35	201			CCGTCAATGA		
	251			TCTGAAATAG		
	301			AGCAGATACT		
	351			TGGGAGAAAA		
	401	AGACTAAGAC	AAATATCGTA	AAAATTGATG	AAAAATTAGA	AGCCGTGGCT
40	451	GATACCGTCG	ACAAGCATGC	CGAAGCATTC	AACGATATCG	CCGATTCATT
	501	GGATGAAACC	AACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG
	551	CCAAACAGAC	GGCCGAAGAA	ACCAAACAAA	ACGTCGATGC	CAAAGTAAAA
	601	GCTGCAGAAA	CTGCAGCAGG	CAAAGCCGAA	GCTGCCGCTG	GCACAGCTAA
4.5	651			AAGCTGTCGC		
45	701			AAAGATAATA		
	751			AGAGTCTGAC		
	801			AAAAATTGGA		
	851			GATACTCGCC		
50	901 951			AACCCGCCAA CTTACAACGT		
50	1001			TCCGAATCGG		
	1051			TGCCGCCAAA		
	1101			CCTACCATGT		
	1151			GCCGCCGACA		
55	1201			CCATAAAGAC		
	1251			AAAACGAGAA		
	1301	GTGCGGAAAA	AACTTATGGA	AACGGTGACA	GCCTCAATAC	GGGCAAATTG
	1351	AAGAACGACA	AGGTCAGCCG	TTTCGACTTT	ATCCGCCAAA	TCGAAGTGGA
	1401			AGAGTGGAGA		
60	1451			TTTCAGACCG		
	1501	CATTCCGGGA	AGATGGTTGC	GAAACGCCAG	TTCAGAATCG	GCGACATAGC
	1551			ACAAGCTTCC		
	1601			TCAGACGATG		
65	1651	ACCATAGATT	TCGCCGCCAA	GCAGGGAAAC	GGCAAAATCG	AACATTTGAA
65	1701			ACCTGGCCGC		
	1751			AGCGGTTCCG		
	1801	AAAGGCAGTT	ACTCCCTCGG	TATCTTTGGC	GGAAAAGCCC	AGGAAGTTGC

WO 01/64922

-68-

PCT/IB01/00452

	1851 1901		GAAGTGAAAA ACTCGAGCAC			ATCGGCCTTG
	1	MATNDDDDVKK	AATVAIAAAY	NNGOEINGFK	AGETIYDIDE	DGTITKKDAT
5	51		GLGLKKVVTN			
	101		DAALDATTNA	-		
	151	DTVDKHAEAF	NDIADSLDET	NTKADEAVKT	ANEAKQTAEE	TKQNVDAKVK
	201	AAETAAGKAE	AAAGTANTAA	DKAEAVAAKV	TDIKADIATN	KDNIAKKANS
	251	ADVYTREESD	SKFVRIDGLN	${\tt ATTEKLDTRL}$	ASAEKSIADH	DTRLNGLDKT
10	301	_	GLAEQAALSG			
	351		AGVAVGTSSG			
	401		KGLQSLTLDQ			
	451		IRQIEVDGQL			
15	501 551		FRIGDIAGEH GKIEHLKSPE			
15	601	~	GKAQEVAGSA			
	001	KGSISDGIFG	GKAQEVAGSA	EAKLANGIKH	IGUAARQUER	mmmm
20	961-983					
20	1		ACGACGACGA			
	51 101		AACAATGGCC CATTGATGAA			
	151		TTGAAGCCGA			
	201		CTGACCAAAA			
25	251		AGCTGCAGAA			
	301		ATGCCGCTTT			
	351		TTGAATAAAT			
	401		AAATATCGTA			
	451	GATACCGTCG	ACAAGCATGC	CGAAGCATTC	AACGATATCG	CCGATTCATT
30	501	GGATGAAACC	AACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG
	551		GGCCGAAGAA			
	601		CTGCAGCAGG			
	651		GACAAGGCCG			
35	701		CGCTACGAAC			
33	751 801		ACACCAGAGA GCTACTACCG			
	851		TGCCGATCAC			
	901		TGCGCAAAGA			
	951		CTGTTCCAAC			
40	1001		CGGCTACAAA			
	1051		CCGAAAACTT			
	1101	TTCGTCCGGT	TCTTCCGCAG	CCTACCATGT	CGGCGTCAAT	TACGAGTGGG
	1151	GATCCGGCGG	AGGCGGCACT	TCTGCGCCCG	ACTTCAATGC	AGGCGGTACC
4 =	1201		GCAACAGCAG			
45	1251		ATCAAGAACG			
	1301		TGACGTTGCG			
	1351		ATCTGCATAC ATCAACCTCA			
	1401 1451		GGTAGGTATC			
50	1501		AACTGTATGG			
20	1551		ACGGCGTATA			
	1601		TGAAGCTTCT			
	1651		CGGATATCCG			
	1701		ATTATTGGCG			
55	1751	GTATTGCGCC	CGATGCGACG	CTACACATAA	TGAATACGAA	TGATGAAACC
	1801	AAGAACGAAA	TGATGGTTGC	AGCCATCCGC	AATGCATGGG	TCAAGCTGGG
	1851		GTGCGCATCG			
	1901		CGACCTTTTC			
60	1951		TCGACTATTC			
60	2001		CAGAGCGATT			
	2051		TTTCATCTTT			
	2101		CCCTATTGCC			
	2151 2201		GCAGGCGTAG			
65	2251		ACCGGGTACA CCATGTGGTG			
55	2301		ACAAACCCGA			
	2351		CGGCACGGCG			
	2,3,4		20001100000		_ CONTOURNIA	JUGGLUGALG

-69-

	2401				ACGACGGCTC	
	2451				GGGACTGCTG	
	2501				TCGGCGACTT	
5	2551				TTCCGTAACG	
3	2601				CCAACTGCAA AAGGCGGTTC	
	2651 2701				GAAACCAAAG	
	2751				GAACAGCGAC	
	2801				ACGAAACCGT	
10	2851				CTGTACACAC	
	2901				CGGCGGCAAG	
	2951	CGGCACGCGG	CAAGGGGGCA	GGCTATCTCA	ACAGTACCGG	ACGACGTGTT
	3001				GATTATTCTT	
	3051	CATCGAAACC	GACGGCGGCC	${\tt TGCTGGCTTC}$	CCTCGACAGC	GTCGAAAAAA
15	3101	CAGCGGGCAG	TGAAGGCGAC	ACGCTGTCCT	ATTATGTCCG	TCGCGGCAAT
	3151	GCGGCACGGA	CTGCTTCGGC	AGCGGCACAT	TCCGCGCCCG	CCGGTCTGAA
	3201				GGAAAACCTG	
	3251				AGACGGTTGA	
20	3301				CCCTACGGCG	
20	3351				CGCCGACGGT	
	3401				ACAGTACCGC	
	3451				TCGGACGGGT CCAACAGGAC	
	3501 3551				GCGGCAGTAC	
25	3601				ACAGCAGCCG	
20	3651				TGCAAATGCA	
	3701				ATGCGGGCGA	
	3751				AAAAACAGCA	
	3801	CACCGGTGCG	GACGAACATG	CGGAAGGCAG	CGTCAACGGC	ACGCTGATGC
30	3851	AGCTGGGCGC	ACTGGGCGGT	GTCAACGTTC	CGTTTGCCGC	AACGGGAGAT
	3901				CTGCTCAAAC	
	3951				CGGCAACAGC	
	4001				TGTCGCAACC	
35	4051				GAACGCGACC	
33	4101				CGCGACTGCA GTCTGGTTGC	
	4151 4201				GGCTTGGCAC	
	4251				CGGACGAGTC	
	4301		CGAGCACCAC			
40						
	1.	MATNDDDVKK	AATVAIAAAY	NNGQEINGFK	AGETIYDIDE	DGTITKKDAT
	51				NVDAKVKAAE	
	101				FAEETKTNIV	
4.54	151				ANEAKQTAEE	
45	201				TDIKADIATN	
	251				ASAEKSIADH	
	301				NVTAAVGGYK YEWGSGGGGT	
	351 401					VTDRDAKINA
50	451				YTGRGVEVGI	
	501				DGGGKDIEAS	
	551				PAGGIAPDAT	
	601	KNEMMVAAIR	NAWVKLGERG	VRIVNNSFGT	TSRAGTADLF	QIANSEEQYR
	651	QALLDYSGGD	KTDEGIRLMQ	QSDYGNLSYH	IRNKNMLFIF	STGNDAQAQP
55	701	NTYALLPFYE	KDAQKGIITV	AGVDRSGEKF	KREMYGEPGT	EPLEYGSNHC
	751	GITAMWCLSA	PYEASVRFTR	TNPIQIAGTS	FSAPIVTGTA	ALLLQKYPWM
	801					SFPFGDFTAD
	851					TIIEGGSLVL
60	901					SGANETVHIK
60	951					GYLNSTGRRV
	1001					TLSYYVRRGN
	1051 1101			-		ATPETVETAA VYADSTAAHA
	1151					GKMRGSTQTV
65	1201	_				IRHDAGDIGY
	1251					VNVPFAATGD
	1301					GLKLSQPLSD
						~

-70-

	1351 1401				ATGKTGARNM GVGYRFLEHH	
5	961c-OR	F46.1				
_	1		ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
	51				CGGTTTCAAA	
	101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACCAAAAA	AGACGCAACT
	151	GCAGCCGATG	TTGAAGCCGA	CGACTTTAAA	GGTCTGGGTC	TGAAAAAAGT
10	201	CGTGACTAAC	CTGACCAAAA	CCGTCAATGA	AAACAAACAA	AACGTCGATG
	251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTTAAC	AACCAAGTTA
	301	GCAGACACTG	${\tt ATGCCGCTTT}$	AGCAGATACT	GATGCCGCTC	TGGATGCAAC
	351	CACCAACGCC	${\tt TTGAATAAAT}$	${\tt TGGGAGAAAA}$	TATAACGACA	TTTGCTGAAG
	401				AAAAATTAGA	
15	451				AACGATATCG	
	501				CGTCAAAACC	
	551				ACGTCGATGC	
	601				GCTGCCGCTG	
20	651				TGCAAAAGTT	
20	701				TTGCTAAAAA	
	751				AGCAAATTTG	
	801				CACACGCTTG	
	851				TGAACGGTTT	
25	901				GGCCTTGCAG	
23	951				GGGTGGATCC	
	1001				GGCAGGTTCT	
	1051				TTCGGCAGCA	
	1101				AAAAATACAA TTAAAGGAAA	
30	1151 1201				GTCCATTCCC	
50	1251				CGGTAGTCCC	
	1301				ACGAACACCA	
	1351				CCCGCTCCCA	
	1401				TGCCCAAAAT	
35	1451				GGCTTGCCGA	
	1501				GGCGACGGAT	
	1551				GGGCAATGCC	
	1601				TCATCGGCGC	
	1651	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGC	ATAAGCGAAG	GCTCAAACAT
40	1701	TGCTGTCATG	CACGGCTTGG	GTCTGCTTTC	CACCGAAAAC	AAGATGGCGC
	1751	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC	TCAAAGACTA	TGCCGCAGCA
	1801	GCCATCCGCG	ATTGGGCAGT	CCAAAACCCC	AATGCCGCAC	AAGGCATAGA
	1851	AGCCGTCAGC	AATATCTTTA	TGGCAGCCAT	CCCCATCAAA	GGGATTGGAG
	1901	CTGTTCGGGG	AAAATACGGC	TTGGGCGGCA	TCACGGCACA	TCCTATCAAG
45	1951				AAAGGGAAAT	
	2001	-			ATACCCGTCC	
	2051				GTTACGGCAA	
	2101				AAAAATGTCA	
50	2151				TGACGGTAAA	
50	2201		GCACGTGAAA	TATGATACGC	TCGAGCACCA	CCACCACCAC
	2251	CACTGA				
	1	MATNDDDVKK	AATVAIAAAY	NNGQEINGFK	AGETIYDIDE	DGTITKKDAT
	51				NVDAKVKAAE	
55	101				FAEETKTNIV	
	151				ANEAKQTAEE	
	201				TDIKADIATN	
	251				ASAEKSIADH	
60	301				GGGGSDLAND	
60	351				SHQLGNLMIQ	
	401				VDGFSLYRIH	
	451				IRLNLTDNRS	
	501 551				AEAFNGTADI	
65	551 601				KMARINDLAD	
05	651				GIGAVRGKYG PYHSRNIRSN	
	701				GFPNFEKHVK	
	701	TUDIALEDING	*/IN A VTHIN TKH	FVIGABLINGK	GE FINE EVUAK	TOTOGRAPH

-71-

PCT/IB01/00452

751 H*

WO 01/64922

	961c-74	1				
5	1		ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
_	51				CGGTTTCAAA	
	101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACCAAAAA	AGACGCAACT
	151				GGTCTGGGTC	
	201	CGTGACTAAC	CTGACCAAAA	CCGTCAATGA	AAACAAACAA	AACGTCGATG
10	251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTTAAC	AACCAAGTTA
	301				GATGCCGCTC	
	351				TATAACGACA	
	401				AAAAATTAGA	
1 ~	451		-		AACGATATCG	
15	501				CGTCAAAACC	
	551	-			ACGTCGATGC	
	601				GCTGCCGCTG	
	651				TGCAAAAGTT	
20	701				TTGCTAAAAA AGCAAATTTG	
20	751 801				CACACGCTTG	
	851				TGAACGGTTT	
	901				GGCCTTGCAG	
	951				GGGTGGATCC	
25	1001				CCGATGCACT	
20	1051				TTGACGCTGG	
	1101				ACAAGGTGCG	
	1151				AATTGAAGAA	
	1201	AGCCGTTTCG	ACTTTATCCG	CCAAATCGAA	GTGGACGGGC	AGCTCATTAC
30	1251	CTTGGAGAGT	GGAGAGTTCC	AAGTATACAA	ACAAAGCCAT	TCCGCCTTAA
	1301	CCGCCTTTCA	GACCGAGCAA	ATACAAGATT	CGGAGCATTC	CGGGAAGATG
	1351	GTTGCGAAAC	GCCAGTTCAG	AATCGGCGAC	ATAGCGGGCG	AACATACATC
	1401	TTTTGACAAG	CTTCCCGAAG	GCGGCAGGGC	GACATATCGC	GGGACGGCGT
	1451				CCTACACCAT	
35	1501				TTGAAATCGC	
	1551				GGATGGAAAA	
	1601				CCGAGAAAGG	
	1651				GTTGCCGGCA	
40	1701				CCTTGCCGCC	AAGCAACTCG
40	1751	AGCACCACCA	CCACCACCAC	TGA		
	1	MATNDDDVKK	AATVAIAAAY	NNGQEINGFK	AGETIYDIDE	DGTITKKDAT
	51	AADVEADDFK	GLGLKKVVTN	LTKTVNENKQ	NVDAKVKAAE	SEIEKLTTKL
	101	ADTDAALADT	DAALDATTNA	LNKLGENITT	FAEETKTNIV	KIDEKLEAVA
45	151				ANEAKQTAEE	~
	201				TDIKADIATN	
	251				ASAEKSIADH	
	301	VSDLRKETRQ	GLAEQAALSG	LFQPYNVGGS	GGGGVAADIG	AGLADALTAP
5 0	351				EKTYGNGDSL	
50	401				SALTAFQTEQ	
	451					GKLTYTIDFA
	501				RHAVISGSVL KOLEHHHHHH	
	551	TGTLGGKAÖE	VAGSAEVKIV	MGTKHTGDAA	KQLERRRRRR	
55						
55	961c-98	13				
	1		ACGACGACGA	AAAAATTOT	GCTGCCACTG	TGGCCATTGC
	51				CGGTTTCAAA	
	101				TTACCAAAAA	
60	151				GGTCTGGGTC	
	201				AAACAAACAA	
	251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTTAAC	AACCAAGTTA
	301	GCAGACACTG	ATGCCGCTTT	AGCAGATACT	GATGCCGCTC	TGGATGCAAC
- · · ·	351				TATAACGACA	
65	401	AGACTAAGAC	AAATATCGTA	AAAATTGATG	AAAAATTAGA	AGCCGTGGCT
	451				AACGATATCG	
	501	GGATGAAACC	AACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG

	551	CCAAACAGAC	GGCCGAAGAA	ACCAAACAAA	ACGTCGATGC	CAAAGTAAAA
	601				GCTGCCGCTG	
	651	TACTGCAGCC	GACAAGGCCG	AAGCTGTCGC	TGCAAAAGTT	ACCGACATCA
	701	AAGCTGATAT	CGCTACGAAC	AAAGATAATA	TTGCTAAAAA	AGCAAACAGT
5	751	GCCGACGTGT	ACACCAGAGA	AGAGTCTGAC	AGCAAATTTG	COMMONOCHO
	801	TGGTCTGAAC	GCTACTACCG	AAAAATTGGA	CACACGCTTG TGAACGGTTT	GCTTCTGCTG
	851	AAAAATCCAT	TGCCGATCAC	AACCCCCCAA	GGCCTTGCAG	ARCARCACC
	901 951	CCTCTCAGACC	TGCGCAAAGA	CTTACAACGT	GGGTGGATCC	GCCGGAGGCG
10	1001	GCICICCGGI	GCCCGACTTC	AATGCAGGCG	GTACCGGTAT	CGGCAGCAAC
10	1051	AGCAGAGCAA	CAACAGCGAA	ATCAGCAGCA	GTATCTTACG	CCGGTATCAA
	1101	GAACGAAATG	TGCAAAGACA	GAAGCATGCT	CTGTGCCGGT	CGGGATGACG
	1151				ATGCCCCCC	
	1201	CATACCGGAG	ACTTTCCAAA	CCCAAATGAC	GCATACAAGA	ATTTGATCAA
15	1251	CCTCAAACCT	GCAATTGAAG	CAGGCTATAC	AGGACGCGGG	GTAGAGGTAG
	1301	GTATCGTCGA	CACAGGCGAA	${\tt TCCGTCGGCA}$	GCATATCCTT	TCCCGAACTG
	1351	TATGGCAGAA	AAGAACACGG	${\tt CTATAACGAA}$	AATTACAAAA	ACTATACGGC
	1401	GTATATGCGG	AAGGAAGCGC	CTGAAGACGG	AGGCGGTAAA	GACATTGAAG
	1451	CTTCTTTCGA	CGATGAGGCC	GTTATAGAGA	CTGAAGCAAA	GCCGACGGAT
20	1501	ATCCGCCACG	TAAAAGAAAT	CGGACACATC	GATTTGGTCT	CCCATATTAT
	1551	TGGCGGGCGT	TCCGTGGACG	GCAGACCTGC	AGGCGGTATT	GCGCCCGATG
	1601				AAACCAAGAA	
	1651	GTTGCAGCCA	TCCGCAATGC	ATGGGTCAAG	CTGGGCGAAC	A CECCCO A CC
25	1701	CATCGTCAAT	AACAGTTTTTG	GAACAACATC	GAGGGCAGGC	CHIRCCHICCACC
25	1751	TTTTCCAAAT	AGCCAATTCG	ACACCACCACT	ACCGCCAAGC ATCCGCCTGA	TCCAACACAC
	1801				TAATAAAAAC	
	1851 1901	TOTAL TACGGC	AGCCDATGAC	GCACAAGCTC	AGCCCAACAC	ATATGCCCTA
	1951				GGCATTATCA	
30	2001				GGAAATGTAT	
	2051	GTACAGAACC	GCTTGAGTAT	GGCTCCAACC	ATTGCGGAAT	TACTGCCATG
	2101	TGGTGCCTGT	CGGCACCCTA	TGAAGCAAGC	GTCCGTTTCA	CCCGTACAAA
	2151	CCCGATTCAA	ATTGCCGGAA	CATCCTTTTC	CGCACCCATC	GTAACCGGCA
	2201	CGGCGGCTCT	GCTGCTGCAG	AAATACCCGT	GGATGAGCAA	CGACAACCTG
35	2251	CGTACCACGT	TGCTGACGAC	GGCTCAGGAC	ATCGGTGCAG	TCGGCGTGGA
	2301	CAGCAAGTTC	GGCTGGGGAC	TGCTGGATGC	GGGTAAGGCC	ATGAACGGAC
	2351				CCGATACGAA	
	2401				TCAGGCACGG	
40	2451	CAAAAAAGGC	GGCAGCCAAC	TGCAACTGCA	CGGCAACAAC	ACCTATACGG
40	2501				TGTTGTACGG CTGATTTATA	
	2551	TCGGATATGC	ACCOMON ACA	CCCACGCCAT	TGTCTATCTG	GCAGATACCG
	2601 2651				TCAAAGGCAG	
	2701				GGCAAACTGC	
45	2751				CATGTCGGCA	
12	2801	GGGCAGGCTA	TCTCAACAGT	ACCGGACGAC	GTGTTCCCTT	CCTGAGTGCC
	2851	GCCAAAATCG	GGCAGGATTA	TTCTTTCTTC	ACAAACATCG	AAACCGACGG
	2901	CGGCCTGCTG	GCTTCCCTCG	ACAGCGTCGA	AAAAACAGCG	GGCAGTGAAG
	2951				GCAATGCGGC	
50	3001	TCGGCAGCGG	CACATTCCGC	GCCCGCCGGT	CTGAAACACG	CCGTAGAACA
	3051	GGGCGGCAGC	AATCTGGAAA	ACCTGATGGT	CGAACTGGAT	GCCTCCGAAT
	3101	CATCCGCAAC	ACCCGAGACG	GTTGAAACTG	CGGCAGCCGA	CCGCACAGAT
	3151	ATGCCGGGCA	TCCGCCCCTA	CGGCGCAACT	TTCCGCGCAG	CGGCAGCCGT
<i>E E</i>	3201	ACAGCATGCG	AATGCCGCCG	ACGGTGTACG	CATCTTCAAC	AGTOTOGOOG
55	3251	CTACCGTCTA	TGCCGACAGT	ACCGCCGCCC	ATGCCGATAT CACAACGGCA	CCCCTCTCCC
	3301	CGCCTGAAAG	CCGTATCGGA	ACCACCCTCC	AACGTGGGAA	CAGGGCCGGTG
	3351	MINCA A CCCA A	A ATTCCCCCCC	AGGACGGIGG	CCGTCGGCAT	TGCCGCGAAA
	3401 3451	7 CCCCCC A A	AMIGCGCGGC	AGIACCCAAA	. CTGGGCATGG	GACGCAGCAC
60	3501	ACCGGCGAAA	AIACGACAGC	ATGCAAAAAC	CGACAGCATT	AGTCTGTTTG
00	3551				GCTATCTCAA	
	3601				CGCAGCACCG	
	3651	ACATGCGGAA	GGCAGCGTCA	ACGGCACGCT	GATGCAGCTG	GGCGCACTGG
	3701	GCGGTGTCA	CGTTCCGTTT	GCCGCAACGG	GAGATTTGAC	GGTCGAAGGC
65	3751	GGTCTGCGCT	ACGACCTGCT	CAAACAGGAT	GCATTCGCCG	AAAAAGGCAG
	3801	TGCTTTGGGC	TGGAGCGGCA	ACAGCCTCAC	: TGAAGGCACG	CTGGTCGGAC
	3851	TCGCGGGTCT	GAAGCTGTCG	CAACCCTTGA	GCGATAAAGC	CGTCCTGTTT

-73-

5	3901 3951 4001 4051 4101 4151	GGGCGGCTTT ATATGCCGCA GGCAACGGCT	ACCGGCGCA CACCCGTCTG GGAACGGCTT CACAGCGGAC	CTGCAGCAAC GTTGCCGGCC GGCACGTTAC	GGACGCGACT CGGCAAGACG TGGGCGCGGA AGCTACGCCG AGGCTACCGG	GGGGCACGCA TGTCGAATTC GTTCCAAACA
10	1 51 101 151 201	AADVEADDFK ADTDAALADT DTVDKHAEAF AAETAAGKAE	GLGLKKVVTN DAALDATTNA NDIADSLDET AAAGTANTAA	LTKTVNENKQ LNKLGENITT NTKADEAVKT DKAEAVAAKV	AGETIYDIDE NVDAKVKAAE FAEETKTNIV ANEAKQTAEE TDIKADIATN	SEIEKLTTKL KIDEKLEAVA TKQNVDAKVK KDNIAKKANS
15	251 301 351 401	VSDLRKETRQ SRATTAKSAA HTGDFPNPND	GLAEQAALSG VSYAGIKNEM AYKNLINLKP	LFQPYNVGGS CKDRSMLCAG AIEAGYTGRG	ASAEKSIADH GGGGTSAPDF RDDVAVTDRD VEVGIVDTGE	NAGGTGIGSN AKINAPPPNL SVGSISFPEL
20	451 501 551 601	IRHVKEIGHI VAAIRNAWVK	DLVSHIIGGR LGERGVRIVN	SVDGRPAGGI NSFGTTSRAG	DIEASFDDEA APDATLHIMN TADLFQIANS MLFIFSTGND	TNDETKNEMM EEQYRQALLD
	651 701 751 801	LPFYEKDAQK WCLSAPYEAS RTTLLTTAQD	GIITVAGVDR VRFTRTNPIQ IGAVGVDSKF	SGEKFKREMY IAGTSFSAPI GWGLLDAGKA	GEPGTEPLEY VTGTAALLLQ MNGPASFPFG TYTGKTIIEG	GSNHCGITAM KYPWMSNDNL DFTADTKGTS
25	851 901 951	SDMRVETKGA DGKGTLYTRL AKIGQDYSFF	LIYNGAASGG GKLLKVDGTA TNIETDGGLL	SLNSDGIVYL IIGGKLYMSA ASLDSVEKTA	ADTDQSGANE RGKGAGYLNS GSEGDTLSYY	TVHIKGSLQL TGRRVPFLSA VRRGNAARTA
30	1001 1051 1101 1151	MPGIRPYGAT RLKAVSDGLD	FRAAAAVQHA HNGTGLRVIA	NAADGVRIFN QTQQDGGTWE	ASESSATPET SLAATVYADS QGGVEGKMRG SLFAGIRHDA	TAAHADMQGR STQTVGIAAK
35	1201 1251 1301 1351	GLRYDLLKQD ATAGVERDLN	AFAEKGSALG GRDYTVTGGF	WSGNSLTEGT	GALGGVNVPF LVGLAGLKLS GARNMPHTRL FLEHHHHHH*	QPLSDKAVLF
	961cL-0	RF46.1	-			
40	1 51 101	ATGAAACACT CTGTAGCGGC CCACTGTGGC	GCACTGGCAG CATTGCTGCT	CCACAAACGA GCCTACAACA	ACAGCCATCC CGACGATGTT ATGGCCAAGA	AAAAAAGCTG AATCAACGGT
45	151 201 251 301	CAAAAAAGAC TGGGTCTGAA	GCAACTGCAG AAAAGTCGTG	CCGATGTTGA ACTAACCTGA	GATGAAGACG AGCCGACGAC CCAAAACCGT GCAGAATCTG	TTTAAAGGTC CAATGAAAAC
	351 401 451	GTTAACAACC CCGCTCTGGA ACGACATTTG	AAGTTAGCAG TGCAACCACC CTGAAGAGAC	ACACTGATGC AACGCCTTGA TAAGACAAAT	CGCTTTAGCA ATAAATTGGG ATCGTAAAAA	GATACTGATG AGAAAATATA TTGATGAAAA
50	501 551 601 651	ATATCGCCGA AAAACCGCCA	TTCATTGGAT ATGAAGCCAA	GAAACCAACA ACAGACGGCC	GCATGCCGAA CTAAGGCAGA GAAGAAACCA AGCAGGCAAA	CGAAGCCGTC AACAAAACGT
55	701 751 801 851	AAAGTTACCG TAAAAAAGCA	ACATCAAAGC AACAGTGCCG	TGATATCGCT ACGTGTACAC	AGGCCGAAGC ACGAACAAAG CAGAGAAGAG CTACCGAAAA	ATAATATTGC TCTGACAGCA
	901 951 1001	CGCTTGGCTT CGGTTTGGAT	CTGCTGAAAA AAAACAGTGT	ATCCATTGCC CAGACCTGCG	GATCACGATA CAAAGAAACC TCCAACCTTA	CTCGCCTGAA CGCCAAGGCC
60	1051 1101 1151 1201	GGTTCTCGAC GCAGCAGGGG	CGTCAGCATT GGAACTTGCC	TCGAACCCGA GAGCGCAGCG	AACGATTCTT CGGGAAATAC GCCATATCGG ATTCAACAGG	CACCTATTCG ATTGGGAAAA
65	1251 1251 1301 1351 1401	AGGAAATATC ATTCCCCCTT AGTCCCGTTG	GGCTACATTG CGACAACCAT ACGGATTTAG	TCCGCTTTTC GCCTCACATT CCTTTACCGC	CGATCACGGG CCGATTCTGA ATCCATTGGG ACAGGGCGGC	CACGAAGTCC TGAAGCCGGT ACGGATACGA
	∵ # ∩ ⊤	**CACCA! CCC	CCCACGGCI	AT GACGGGCC	ನಿರಾರಾಭಾವನ	COCIMICCOG

-74-

PCT/IB01/00452

	1451 1501				ACGACATAAA CGCAGCACCG	
	1551				GCTGACGCAA	
	1601				CCGAGCTGGA	
5	1651				GATATCGTTA	
	1701				CGATGCCGTG	
	1751				GCTTGGGTCT	
	1801				GCAGATATGG	
10	1851				GGCAGTCCAA	
10	1901 1951				TCTTTATGGC TACGGCTTGG	
	2001				CGCGATCGCA	
	2051				ATGCGGCATA	
	2101				TCAAACTTGG	
15	2151				GCCGCCGTCA	
	2201				AGACAGGCGT	
	2251		TTCCGAATTT	TGAGAAGCAC	GTGAAATATG	ATACGTAACT
	2301	CGAG				
20	1	MKABDGKIT.T	ምል ፲፻.አምፑሮዌር	ערורוואיזייע עיז ע	KKAATVAIAA	AVNNCOFTNO
20	51				FKGLGLKKVV	
	101				DTDAALDATT	
	151				AFNDIADSLD	
	201	KTANEAKQTA	EETKQNVDAK	VKAAETAAGK	AEAAAGTANT	AADKAEAVAA
25	251				SDSKFVRIDG	
	301				RQGLAEQAAL	~
	351				HLFGSRGELA	
	401				HEVHSPFDNH	
30	451				GYPAPKGARD	
30	501 551				GVGDGFKRAT QGISEGSNIA	
	601				NPNAAQGIEA	
	651				LPKGKSAVSD	
	701				NGKNVKLADQ	
35	751	GKGFPNFEKH			~	
	•					
	961cL-7					
40	1				ACAGCCATCC	
40	51				CGACGATGTT ATGGCCAAGA	
	101 151				GATGAAGACG	
	201				AGCCGACGAC	
	251				CCAAAACCGT	
45	301		-		GCAGAATCTG	
	351	GTTAACAACC	AAGTTAGCAG	ACACTGATGC	CGCTTTAGCA	GATACTGATG
	401	CCGCTCTGGA	TGCAACCACC	AACGCCTTGA	ATAAATTGGG	AGAAAATATA
	451				ATCGTAAAAA	
50	501				GCATGCCGAA	
50	551				CTAAGGCAGA	
	601 651				GAAGAAACCA	
	701				AGCAGGCAAA AGGCCGAAGC	
	751				ACGAACAAAG	
55	801				CAGAGAAGAG	
	851				CTACCGAAAA	
	901	CGCTTGGCTT	CTGCTGAAAA	ATCCATTGCC	GATCACGATA	CTCGCCTGAA
	951	CGGTTTGGAT	AAAACAGTGT	CAGACCTGCG	CAAAGAAACC	CGCCAAGGCC
60	1001	TTGCAGAACA	AGCCGCGCTC	TCCGGTCTGT	TCCAACCTTA	CAACGTGGGT
60	1051				ATCGGTGCGG	
	1101				CAAAGGTTTG	
	1151				AACTGAAGCT	
	1201				AGCCTCAATA	
65	1251 1301				TATCCGCCAA AGTTCCAAGT	
	1351				GAGCAAATAC	
	1401			CGAAACGCCA		
	7.407	GCMITCCCCC	THIOHICOILG	COLHILLCCCCCT	GIICHGENEIC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

	1451				CCGAAGGCGG	
	1501 1551				GCCGGCGAA CGGCAAAATC	
	1601				CCGCCGATAT	
5	1651				GTCCTTTACA	
J	1701				CGGAAAAGCC	
	1751				GCATACGCCA	
	1801			CCACCACCAC		,
		**				
10	1	MKHFPSKVLT	TAILATFCSG	ALAATNDDDV	KKAATVAIAA	AYNNGQEING
	51	FKAGETIYDI	DEDGTITKKD	ATAADVEADD	${\tt FKGLGLKKVV}$	TNLTKTVNEN
	101	KQNVDAKVKA	AESEIEKLTT	KLADTDAALA	$\mathtt{DTDAALDATT}$	NALNKLGENI
	151	TTFAEETKTN	IVKIDEKLEA	VADTVDKHAE	AFNDIADSLD	ETNTKADEAV
. .	201				AEAAAGTANT	
15	251				SDSKFVRIDG	
	301				RQGLAEQAAL	
	351				QSLTLDQSVR	
	401				IEVDGQLITL	
20	451				GDIAGEHTSF EHLKSPELNV	
20	501 551				OEVAGSAEVK	
	601	AAKOLEHHHH	~	ANDUTEDICE	QEVAG5AEVK	TANGTERITOR
	001	1111/2111111111	****			
25	061-7-0	0.2	,			
23	961cL-9		ጥጥር C አጥር C አ አ	አርሞአርሞርአርር	ACAGCCATCC	ጣጥርርር እርጥጥጥ
	51				CGACGATGTT	
	101				ATGGCCAAGA	
	151				GATGAAGACG	
30	201	CAAAAAAGAC	GCAACTGCAG	CCGATGTTGA	AGCCGACGAC	TTTAAAGGTC
	251	TGGGTCTGAA	AAAAGTCGTG	ACTAACCTGA	CCAAAACCGT	CAATGAAAAC
	301	AAACAAAACG	TCGATGCCAA	AGTAAAAGCT	GCAGAATCTG	AAATAGAAAA
	351				CGCTTTAGCA	
25	401				ATAAATTGGG	
35	451				ATCGTAAAAA	
	501				GCATGCCGAA	
	551				CTAAGGCAGA	
	601				GAAGAAACCA AGCAGGCAAA	
40	651 701				AGGCCGAAGC	
40	751				ACGAACAAAG	
	801				CAGAGAAGAG	
	851				CTACCGAAAA	
	901				GATCACGATA	
45	951	CGGTTTGGAT	AAAACAGTGT	CAGACCTGCG	CAAAGAAACC	CGCCAAGGCC
	1001	TTGCAGAACA	AGCCGCGCTC	TCCGGTCTGT	TCCAACCTTA	CAACGTGGGT
	1051	GGATCCGGCG	GAGGCGGCAC	TTCTGCGCCC	GACTTCAATG	CAGGCGGTAC
	1101				AGCGAAATCA	
70	1151				AAGACAGAAG	
50	1201				AGGGATGCCA	
	1251				TCCAAACCCA	
	1301				TTGAAGCAGG GGCGAATCCG	
	1351 1401				ACACGGCTAT	
55	1451				AAGCGCCTGA	
	1501				GAGGCCGTTA	
	1551				AGAAATCGGA	
	1601				TGGACGGCAG	
	1651				ATGAATACGA	
60	1701	CAAGAACGAA	ATGATGGTTG	CAGCCATCCG	CAATGCATGG	GTCAAGCTGG
	1751				GTTTTGGAAC	
	1801				AATTCGGAGG	
	1851				TAAAACAGAC	
65	1901				TGTCCTACCA	
65	1951				AATGACGCAC	
	2001				AAAAGACGCT	
	2051	TTATCACAGT	CGCAGGCGTA	GACCGCAGTG	GAGAAAAGTT	CAAACGGGAA

	2101	ATGTATGGAG	AACCGGGTAC	AGAACCGCTT	GAGTATGGCT	CCAACCATTG
	2151	CGGAATTACT	GCCATGTGGT	GCCTGTCGGC	ACCCTATGAA	GCAAGCGTCC
	2201	GTTTCACCCG	TACAAACCCG	ATTCAAATTG	CCGGAACATC	CTTTTCCGCA
_	2251	CCCATCGTAA	CCGGCACGGC	GGCTCTGCTG	CTGCAGAAAT	ACCCGTGGAT
5	2301	GAGCAACGAC	AACCTGCGTA	CCACGTTGCT	GACGACGGCT	CAGGACATCG
	2351	GTGCAGTCGG	CGTGGACAGC	AAGTTCGGCT	GGGGACTGCT	$\tt GGATGCGGGT$
	2401	AAGGCCATGA	ACGGACCCGC	GTCCTTTCCG	TTCGGCGACT	TTACCGCCGA
	2451	TACGAAAGGT	ACATCCGATA	TTGCCTACTC	CTTCCGTAAC	GACATTTCAG
4.0	2501	GCACGGGCGG	CCTGATCAAA	AAAGGCGGCA	GCCAACTGCA	ACTGCACGGC
10	2551	AACAACACCT	ATACGGGCAA	AACCATTATC	GAAGGCGGTT	CGCTGGTGTT
	2601	GTACGGCAAC	AACAAATCGG	ATATGCGCGT	CGAAACCAAA	GGTGCGCTGA
	2651				TGAACAGCGA	
	2701				AACGAAACCG	
1.5	2751				GCTGTACACA	
15	2801				TCGGCGGCAA	
	2851				AACAGTACCG	
	2901				GGATTATTCT	
	2951				CCCTCGACAG	
20	3001				TATTATGTCC	
20	3051				TTCCGCGCCC	
	3101				TGGAAAACCT	
	3151				GAGACGGTTG	
	3201				CCCCTACGGC	
25	3251				CCGCCGACGG	
23	3301				GACAGTACCG	
	3351				ATCGGACGGG	
	3401				CCCAACAGGA	
	3451				CGCGGCAGTA	
30	3501 3551				GACAGCAGCC GTGCAAATGC	
30	3601				GATGCGGGCG	
	3651				CAAAAACAGC	
	3701				GCGTCAACGG	
	3751				CCGTTTGCCG	
35	3801				CCTGCTCAAA	
	3851				GCGGCAACAG	
	3901				CTGTCGCAAC	
	3951				GGAACGCGAC	
	4001				GCGCGACTGC	
40	4051				CGTCTGGTTG	
	4101	CGCGGATGTC	GAATTCGGCA	ACGGCTGGAA	CGGCTTGGCA	CGTTACAGCT
	4151	ACGCCGGTTC	CAAACAGTAC	GGCAACCACA	GCGGACGAGT	CGGCGTAGGC
	4201	TACCGGTTCT	GACTCGAG			
45	1	MKHFPSKVLT	TAILATFCSG	ALAATNDDDV	KKAATVAIAA	AYNNGQEING
	51	FKAGETIYDI	DEDGTITKKD	ATAADVEADD	FKGLGLKKVV	TNLTKTVNEN
	101	KQNVDAKVKA	AESEIEKLTT	KLADTDAALA	DTDAALDATT	NALNKLGENI
	151	TTFAEETKTN	IVKIDEKLEA	VADTVDKHAE	AFNDIADSLD	ETNTKADEAV
70	201				AEAAAGTANT	
50	251				SDSKFVRIDG	
	301				RQGLAEQAAL	
	351				AAVSYAGIKN	
	401				NDAYKNLINL	
<i>= =</i>	451				NENYKNYTAY	
55	501				HIDLVSHIIG	
	551				VKLGERGVRI	
	601				EGIRLMQQSD	
	651				QKGIITVAGV	
60	701				ASVRFTRTNP	
60	751				QDIGAVGVDS	
	801				DISGTGGLIK	~ ~
	851				GALIYNGAAS	
	901				RLGKLLKVDG	
65	951				FFTNIETDGG	
05	1001				AGLKHAVEQG	
	1051				ATFRAAAAVQ	
	1101	AYVTAALGMA	DMUAHAATEU	GKKLKAVSDG	LDHNGTGLRV	TWÖ.LÖÖDGG.L

-77-

1151	WEQGGVEGKM	RGSTQTVGIA	AKTGENTTAA	ATLGMGRSTW	SENSANAKTD
1201	SISLFAGIRH	DAGDIGYLKG	LFSYGRYKNS	ISRSTGADEH	AEGSVNGTLM
1251	QLGALGGVNV	PFAATGDLTV	EGGLRYDLLK	QDAFAEKGSA	LGWSGNSLTE
1301	GTLVGLAGLK	LSQPLSDKAV	LFATAGVERD	LNGRDYTVTG	GFTGATAATG
1351	KTGARNMPHT	RLVAGLGADV	EFGNGWNGLA	RYSYAGSKQY	GNHSGRVGVG
1401	YRF*				

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention. For instance, the use of proteins from other strains is envisaged [e.g. see WO00/66741 for polymorphic sequences for ORF4, ORF40, ORF46, 225, 235, 287, 519, 726, 919 and 953].

EXPERIMENTAL DETAILS

FPLC protein purification

5

10

20

15 The following table summarises the FPLC protein purification that was used:

Protein	PI	Column	Buffer	pН	Protocol
121.1 ^{untagged}	6.23	Mono Q	Tris	8.0	A
128.1 ^{untagged}	5.04	Mono Q	Bis-Tris propane	6.5	A
406.1L	7.75	Mono Q	Diethanolamine	9.0	В
576.1L	5.63	Mono Q	Tris	7.5	В
593 ^{untagged}	8.79	Mono S	Hepes	7.4	A
726 ^{untagged}	4.95	Hi-trap S	Bis-Tris	6.0	A
919 ^{untagged}	10.5(-leader)	Mono S	Bicine	8.5	С
919Lorf4	10.4(-leader)	Mono S	Tris	8.0	В
920L	6.92(-leader)	Mono Q	Diethanolamine	8.5	A
953L	7.56(-leader)	Mono S	MES	6.6	D
982 ^{untagged}	4.73	Mono Q	Bis-Tris propane	6.5	A
919-287	6.58	Hi-trap Q	Tris	8.0	A
953-287	4.92	Mono Q	Bis-Tris propane	6.2	A

Buffer solutions included 20-120 mM NaCl, 5.0 mg/ml CHAPS and 10% v/v glycerol. The dialysate was centrifuged at 13000g for 20 min and applied to either a mono Q or mono S FPLC ion-exchange resin. Buffer and ion exchange resins were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual [Pharmacia: FPLC Ion Exchange and Chromatofocussing; Principles and Methods. Pharmacia

-78-

Publication]. Proteins were eluted using a step-wise NaCl gradient. Purification was analysed by SDS-PAGE and protein concentration determined by the Bradford method.

The letter in the 'protocol' column refers to the following:

5

10

15

20

25

30

FPLC-A: Clones 121.1, 128.1, 593, 726, 982, periplasmic protein 920L and hybrid proteins 919-287, 953-287 were purified from the soluble fraction of E.coli obtained after disruption of the cells. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD₅₅₀ reached 0.6-08. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C. When necessary cells were stored at -20°C. All subsequent procedures were performed on ice or at 4°C. For cytosolic proteins (121.1, 128.1, 593, 726 and 982) and periplasmic protein 920L, bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim). Cells were lysed by sonication using a Branson Sonifier 450. Disrupted cells were centrifuged at 8000g for 30 min to sediment unbroken cells and inclusion bodies and the supernatant taken to 35% v/v saturation by the addition of 3.9 M (NH₄)₂SO₄. The precipitate was sedimented at 8000g for 30 minutes. The supernatant was taken to 70% v/v saturation by the addition of 3.9 M (NH₄)₂SO₄ and the precipitate collected as above. Pellets containing the protein of interest were identified by SDS-PAGE and dialysed against the appropriate ion-exchange buffer (see below) for 6 hours or overnight. The periplasmic fraction from E.coli expressing 953L was prepared according to the protocol of Evans et. al. [Infect.Immun. (1974) 10:1010-1017] and dialysed against the appropriate ion-exchange buffer. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual (Pharmacia). Buffer solutions included 20 mM NaCl, and 10% (v/v) glycerol. The dialysate was centrifuged at 13000g for 20 min and applied to either a mono Q or mono S FPLC ionexchange resin. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual (Pharmacia). Proteins were eluted from the ion-exchange resin using either step-wise or continuous NaCl gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method. Cleavage of the leader peptide of periplasmic proteins was demonstrated by sequencing the NH₂-terminus (see below).

5

10

15

20

25

30

-79-

FPLC-B: These proteins were purified from the membrane fraction of E.coli. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium. Clones 406.1L and 919LOrf4 were grown at 30°C and Orf25L and 576.1L at 37°C until the OD₅₅₀ reached 0.6-0.8. In the case of 919LOrf4, growth at 30°C was essential since expression of recombinant protein at 37°C resulted in lysis of the cells. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed at 4°C. Bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim) and lysed by osmotic shock with 2-3 passages through a French Press. Unbroken cells were removed by centrifugation at 5000g for 15 min and membranes precipitated by centrifugation at 100000g (Beckman Ti50, 38000rpm) for 45 minutes. A Dounce homogenizer was used to re-suspend the membrane pellet in 7.5 ml of 20 mM Tris-HCl (pH 8.0), 1.0 M NaCl and complete protease inhibitor. The suspension was mixed for 2-4 hours, centrifuged at 100000g for 45 min and the pellet resuspended in 7.5 ml of 20mM Tris-HCl (pH 8.0), 1.0M NaCl, 5.0mg/ml CHAPS, 10% (v/v) glycerol and complete protease inhibitor. The solution was mixed overnight, centrifuged at 100000g for 45 minutes and the supernatant dialysed for 6 hours against an appropriately selected buffer. In the case of Orf25.L, the pellet obtained after CHAPS extraction was found to contain the recombinant protein. This fraction, without further purification, was used to immunise mice.

FPLC-C: Identical to FPLC-A, but purification was from the soluble fraction obtained after permeabilising *E.coli* with polymyxin B, rather than after cell disruption.

FPLC-D: A single colony harbouring the plasmid of interest was grown overnight at 37°C in 20 ml of LB/Amp (100 μg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at 30°C until the OD₅₅₀ reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed on ice or at 4°C. Cells were resuspended in 20mM Bicine (pH 8.5), 20mM NaCl, 10% (v/v) glycerol, complete protease inhibitor (Boehringer-Mannheim) and disrupted using a Branson Sonifier 450. The sonicate was centrifuged at 8000g for 30 min to sediment unbroken cells and

5

10

inclusion bodies. The recombinant protein was precipitated from solution between 35% v/v and 70% v/v saturation by the addition of 3.9M $(NH_4)_2SO_4$. The precipitate was sedimented at 8000g for 30 minutes, resuspended in 20 mM Bicine (pH 8.5), 20 mM NaCl, 10% (v/v) glycerol and dialysed against this buffer for 6 hours or overnight. The dialysate was centrifuged at 13000g for 20 min and applied to the FPLC resin. The protein was eluted from the column using a step-wise NaCl gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method.

Cloning strategy and oligonucleotide design

Genes coding for antigens of interest were amplified by PCR, using oligonucleotides designed on the basis of the genomic sequence of *N. meningitidis* B MC58. Genomic DNA from strain 2996 was always used as a template in PCR reactions, unless otherwise specified, and the amplified fragments were cloned in the expression vector pET21b+ (Novagen) to express the protein as C-terminal His-tagged product, or in pET-24b+(Novagen) to express the protein in 'untagged' form (*e.g.* ΔG 287K).

Where a protein was expressed without a fusion partner and with its own leader peptide (if present), amplification of the open reading frame (ATG to STOP codons) was performed.

Where a protein was expressed in 'untagged' form, the leader peptide was omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

The melting temperature of the primers used in PCR depended on the number and type of hybridising nucleotides in the whole primer, and was determined using the formulae:

$$T_{m1} = 4 (G+C)+2 (A+T)$$
 (tail excluded)

$$T_{m2} = 64.9 + 0.41 \text{ (\% GC)} - 600/N$$
 (whole primer)

The melting temperatures of the selected oligonucleotides were usually 65-70°C for the whole oligo and 50-60°C for the hybridising region alone.

Oligonucleotides were synthesised using a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2.0ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were centrifuged and the pellets resuspended in water.

		Sequences	Restriction site
Orf1L	Fwd	CGCGGATCCGCTAGC-AAAACAACCGACAAACGG	NheI
	Rev	CCCG <u>CTCGAG</u> -TTACCAGCGGTAGCCTA	XhoI
Orf1	Fwd	CTAGCTAGC-GGACACACTTATTTCGGCATC	NheI
	Rev	CCCGCTCGAG- TTACCAGCGGTAGCCTAATTTG	XhoI
Orf1LOmpA	Fwd		NdeI-(NheI)
_	Rev	CCCGCTCGAG-	XhoI
Orf4L	Fwd	CGCGGATCCCATATG-AAAACCTTCTTCAAAACC	NdeI
	Rev	CCCGCTCGAG-TTATTTGGCTGCGCCTTC	XhoI
Orf7-1L	Fwd	GCGGC <u>ATTAAT</u> -ATGTTGAGAAAATTGTTGAAATGG	AseI
	Rev	GCGGC <u>CTCGAG</u> -TTATTTTTCAAAATATATTTGC	XhoI
Orf9-1L	Fwd	GCGGC <u>CATATG</u> -TTACCTAACCGTTTCAAAATGT	NdeI
011) 112	Rev	GCGGC <u>CTCGAG-TTATTTCCGAGGTTTTCGGG</u>	XhoI
Orf23L	Fwd	CGCGGATCC <u>CATATG</u> -ACACGCTTCAAATATTC	NdeI
011231	Rev	CCCGCTCGAG-TTATTTAAACCGATAGGTAAA	XhoI
Orf25-1 His	Fwd	CGCGGATCCCATATG-GGCAGGGAAGAACCGC	NdeI
OF125-1 THS	Rev	GCCCAAGCTT-ATCGATGGAATAGCCGCG	HindIII
O .000 1 b III'			
Orf29-1 b-His	Fwd	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCG	NheI
(MC58)	Rev	CCCGCTCGAG-TTTGTCTAAGTTCCTGATAT	XhoI
Orf29-1 b-L	Fwd	CCCG <u>CTCGAG</u> -ATTCCCACCTGCCATC CGCGGATCC <u>GCTAGC</u> -ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTAATTCCCACCTGCCATC	XhoI
(MC58)			
Orf29-1 c-His	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
(MC58)	Rev	CCCGCTCGAG-TTGGACGATGCCCGCGA	XhoI
Orf29-1 c-L	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
(MC58)	Rev	CCCGCTCGAG-TTATTGGACGATGCCCGC	XhoI
Orf25L	Fwd	CGCGGATCC <u>CATATG</u> -TATCGCAAACTGATTGC	NdeI
	Rev	CCCG <u>CTCGAG</u> -CTAATCGATGGAATAGCC	XhoI
Orf37L	Fwd	CGCGGATCC <u>CATATG</u> -AAACAGACAGTCAAATG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAATAACCCGCCTTCAG	XhoI
Orf38L	Fwd	CGCGGATCC <u>CATATG</u> - TTACGTTTGACTGCTTTAGCCGTATGCACC	NdeI
	Rev	CCCG <u>CTCGAG</u> - TTATTTTGCCGCGTTAAAAGCGTCGGCAAC	XhoI
Orf40L	Fwd	CGCGGATCC <u>CATATG</u> -AACAAAATATACCGCAT	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCACTGATAACCGAC	XhoI
Orf40.2-His	Fwd	CGCGGATCC <u>CATATG</u> -ACCGATGACGACGATTTAT	NdeI
	Rev	GCCCAAGCTT-CCACTGATAACCGACAGA	HindIII
Orf40.2L	Fwd	CGCGGATCCCATATG-AACAAAATATACCGCAT	NdeI
	Rev	GCCCAAGCTT-TTACCACTGATAACCGAC	HindIII
Orf46-2L	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTATTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf46-2	Fwd	GGGAATTCCATATG-TCAGATTTGGCAAACGATTCTT	NdeI
OKITO M	Rev	CCCGCTCGAG-TTATTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf46.1L	Fwd		NdeI

	Rev	CCCG <u>CTCGAG</u> -TTACGTATCATATTTCACGTGC	XhoI
orf46. (His-GST)	Fwd	GGGAATTC <u>CATATG</u> CACGTGAAATATGATACGAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
orf46.1-His	Fwd	GGGAATTC <u>CATATG</u> TCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGCGTATCATATTTCACGTGC	XhoI
orf46.2-His	Fwd	GGGAATTC <u>CATATG</u> TCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf65-1-(His/GST)	Fwd	CGCGGATCCCATATG-CAAAATGCGTTCAAAATCCC	BamHI-NdeI
(MC58)	Rev	CGCGGATCC <u>CATATG</u> -AACAAAATATACCGCAT	XhoI
		CCCG <u>CTCGAG</u> -TTTGCTTTCGATAGAACGG	
Orf72-1L	Fwd	GCGGC <u>CATATG</u> -GTCATAAAATATACAAATTTGAA	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTAGCCTGAGACCTTTGCAAATT	XhoI
Orf76-1L	Fwd	GCGGC <u>CATATG</u> -AAACAGAAAAAAACCGCTG	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTACGGTTTGACACCGTTTTC	XhoI
Orf83.1L	Fwd	CGCGGATCC <u>CATATG</u> -AAAACCCTGCTCCTC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATCCTCCTTTGCGGC	XhoI
Orf85-2L	Fwd	GCGGC <u>CATATG</u> -GCAAAAATGATGAAATGGG	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTATCGGCGCGGGCCC	XhoI
Orf91L (MC58)	Fwd	GCGGCCATATGAAAAAATCCTCCCTCATCA	NdeI
	Rev	GCGGCCTCGAGTTATTTGCCGCCGTTTTTGGC	XhoI
Orf91-His(MC58)	Fwd	GCGGCCATATGGCCCCTGCCGACGCGGTAAG	NdeI
	Rev	GCGGCCTCGAGTTTGCCGCCGTTTTTGGCTTTC	XhoI
Orf97-1L	Fwd	GCGGC <u>CATATG</u> -AAACACATACTCCCCCTGA	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTATTCGCCTACGGTTTTTTG	XhoI
Orf119L (MC58)	Fwd	GCGGCCATATGATTTACATCGTACTGTTTC	NdeI
	Rev	GCGGCCTCGAGTTAGGAGAACAGGCGCAATGC	XhoI
Orf119-His(MC58)	Fwd	GCGGCCATATGTACAACATGTATCAGGAAAAC	NdeI
	Rev	GCGGCCTCGAGGGAGAACAGGCGCAATGCGG	XhoI
Orf137.1 (His- GST) (MC58)	Fwd	CGC <u>GGATCCGCTAGC</u> TGCGGCACGGCGGG	BamHI-NheI
	Rec	CCCG <u>CTCGAG</u> ATAACGGTATGCCGCCAG	XhoI
Orf143-1L	Fwd	CGCGGATCC <u>CATATG</u> -GAATCAACACTTTCAC	NdeI
	Rev	CCCG <u>CTCGAG-</u> TTACACGCGGTTGCTGT	XhoI
008	Fwd	CGCGGATCC <u>CATATG</u> -AACAACAGACATTTTG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCTGTCCGGTAAAAG	XhoI
050-1(48)	Fwd	CGCGGATCC <u>GCTAGC</u> -ACCGTCATCAAACAGGAA	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAAGATTCGACGGGGA	XhoI
105	Fwd	CGCGGATCC <u>CATATG</u> -TCCGCAAACGAATACG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAGTGTTCTGCCAGTTT	XhoI
111L	Fwd	CGCGGATCC <u>CATATG</u> -CCGTCTGAAACACG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTAGCGGAGCAGTTTTTC	XhoI
117-1	Fwd	CGCGGATCC <u>CATATG</u> -ACCGCCATCAGCC	NdeI
	Rev	CCCGCTCGAG-TTAAAGCCGGGTAACGC	XhoI
121-1	Fwd	GCGGC <u>CATATG</u> -GAAACACAGCTTTACATCGG	NdeI
	Rev	GCGGCCTCGAG-TCAATAATAATATCCCGCG	XhoI

PCT/IB01/00452

122-1	Fwd	GCGGC <u>CATATG</u> -ATTAAAATCCGCAATATCC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTAAATCTTGGTAGATTGGATTTGG	XhoI
128-1	Fwd	GCGGC <u>CATATG</u> -ACTGACAACGCACTGCTCC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAGACCGCGTTGTCGAAAC	XhoI
148	Fwd	CGCGGATCC <u>CATATG</u> -GCGTTAAAAACATCAAA	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAGCCCTTCATACAGC	XhoI
149.1L (MC58)	Fwd	GCGGCATTAATGGCACAAACTACACTCAAACC	AseI
	Rev	GCGGCCTCGAGTTAAAACTTCACGTTCACGCCG	XhoI
149.1-His(MC58)	Fwd	GCGGCATTAATGCATGAAACTGAGCAATCGGTGG	AseI
	Rev	GCGGCCTCGAGAAACTTCACGTTCACGCCGCCGGTAAA	XhoI
205 (His-GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> GGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGATAATGGCGGCGGCGG	XhoI
206L	Fwd	CGCGGATCC <u>CATATG</u> -TTTCCCCCCGACAA	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCATTCTGTAAAAAAAGTATG	XhoI
214 (His-GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> CTTCAAAGCGACAGCAG	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> TTCGGATTTTTGCGTACTC	XhoI
216	Fwd	CGCGGATCC <u>CATATG</u> -GCAATGGCAGAAAACG	NdeI
	Rev	CCCG <u>CTCGAG</u> -CTATACAATCCGTGCCG	XhoI
225-1L	Fwd	CGCGGATCC <u>CATATG</u> -GATTCTTTTTCAAACC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAGTTCAGAAAGCGGG	XhoI
235L	Fwd	CGCGGATCC <u>CATATG</u> -AAACCTTTGATTTTAGG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTTGGGCTGCTCTTC	XhoI
243	Fwd	CGCGGATCC <u>CATATG</u> -GTAATCGTCTGGTTG	NdeI
	Rev	CCCG <u>CTCGAG</u> -CTACGACTTGGTTACCG	XhoI
247-1L	Fwd	GCGGC <u>CATATG</u> -AGACGTAAAATGCTAAAGCTAC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAAAGTGTTCTGTTTGCGC	XhoI
264-His	Fwd	GCCGC <u>CATATG</u> -TTGACTTTAACCCGAAAAA	NdeI
	Rev	GCCGC <u>CTCGAG</u> -GCCGGCGGTCAATACCGCCCGAA	XhoI
270 (His-GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> GCGCAATGCGATTTGAC	BamHI-NdeI
	Rev	CCCGCTCGAGTTCGGCGGTAAATGCCG	XhoI
274L	Fwd	GCGGC <u>CATATG</u> -GCGGGGCCGATTTTTGT	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTATTTGCTTTCAGTATTATTG	XhoI
283L	Fwd	GCGGC <u>CATATG</u> -AACTTTGCTTTATCCGTCA	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTAACGGCAGTATTTGTTTAC	XhoI
285-His	Fwd	CGC <u>GGATCC</u> CATATGGGTTTGCGCTTCGGGC	BamHI
	Rev	GCCCAAGCTTTTTCCTTTGCCGTTTCCG	HindIII
286-His	Fwd	CGCGGATCC <u>CATATG</u> -GCCGACCTTTCCGAAAA	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -GAAGCGCGTTCCCAAGC	XhoI
286L	Fwd	CGCGGATCC <u>CATATG</u> -CACGACACCCGTAC	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTAGAAGCGCGTTCCCAA	XhoI
287L	Fwd	CTA <u>GCTAGC</u> -TTTAAACGCAGCGTAATCGCAATGG	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC	XhoI

PCT/IB01/00452

287	Fwd	CTA <u>GCTAGC</u> -GGGGGCGGCGGTGGCG	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC	XhoI
287LOrf4	Fwd	CTA <u>GCTAGC</u> GCTCATCCTCGCCGCC- TGCGGGGGCGGCGGT	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC	XhoI
287-fu	Fwd	CGG <u>GGATCC</u> -GGGGGCGCGGTGGCG	BamHI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC	XhoI
287-His	Fwd	CTA <u>GCTAGC</u> -GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC*	XhoI
287-His(2996)	Fwd	CTAGCTAGC-TGCGGGGGGGGGGGGGGGG	NheI
	Rev	CCCG <u>CTCGAG</u> -ATCCTGCTCTTTTTTGCC	XhoI
Δ1 287-His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC §	NheI
Δ2 287-His	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGGCAGT§	NheI
Δ3 287-His	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA §	NheI
Δ4 287-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG §	NheI
Δ4 287MC58-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG§	NheI
287a-His	Fwd	CGC <u>CATATG</u> -TTTAAACGCAGCGTAATCGC	NdeI
	Rev	CCCGCTCGAG-AAAATTGCTACCGCCATTCGCAGG	XhoI
287b-His	Fwd	CGCCATATG-GGAAGGGTTGATTTGGCTAATGG	NdeI
287b-2996-His	Rev	CCCGCTCGAG-CTTGTCTTTATAAATGATGACATATTTG	XhoI
287b-MC58-His	Rev	CCCGCTCGAG-TTTATAAAAGATAATATATTGATTGATTCC	XhoI
287c-2996-His	Fwd	CGCGCTAGC-ATGCCGCTGATTCCCGTCAATC §	NheI
'287 ^{untagged} ',(2996)	Fwd	CTAGCTAGC-GGGGGGGGGGGGGGGG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG287-His *	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC	XhoI
ΔG287K(2996)	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
, ,	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-L	Fwd	CGCGGATCC <u>GCTAGC</u> - TTTGAACGCAGTGTGATTGCAATGGCTTGTATTTTTGCC CTTTCAGCCTGT TCGCCCGATGTTAAATCGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-Orf4L	Fwd	CGCGGATCC <u>GCTAGC</u> - AAAACCTTCTTCAAAACCCTTTCCGCCGCCGCACTCGCG CTCATCCTCGCCGCCTGC TCGCCCGATGTTAAATCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
292L	Fwd	CGCGGATCC <u>CATATG</u> -AAAACCAAGTTAATCAAA	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTGATTTTTGCGGATGA	XhoI
308-1	Fwd	CGCGGATCC <u>CATATG</u> -TTAAATCGGGTATTTTATC	NdeI
	Rev	CCCGCTCGAG-TTAATCCGCCATTCCCTG	XhoI
401L	Fwd	GCGGC <u>CATATG</u> -AAATTACAACAATTGGCTG	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTACCTTACGTTTTCAAAG	XhoI
406L	Fwd	CGCGGATCCCATATG-CAAGCACGGCTGCT	NdeI
	Rev	CCCGCTCGAG-TCAAGGTTGTCCTTGTCTA	XhoI
502-1L	Fwd	CGCGGATCCCATATG-ATGAAACCGCACAAC	NdeI
·	Rev	CCCGCTCGAG-TCAGTTGCTCAACACGTC	XhoI

502-A (His-GST)	Fwd	CGC <u>GGATCCCATATG</u> GTAGACGCGCTTAAGCA	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> AGCTGCATGGCGGCG	XhoI
503-1L	Fwd	CGCGGATCC <u>CATATG</u> -GCACGGTCGTTATAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -CTACCGCGCATTCCTG	XhoI
519-1L	Fwd	GCGGCCATATG-GAATTTTTCATTATCTTGTT	NdeI
	Rev	GCGGCCTCGAG-TTATTTGGCGGTTTTGCTGC	XhoI
525-1L	Fwd	GCGGCCATATG-AAGTATGTCCGGTTATTTTC	NdeI
	Rev	GCGGCCTCGAG-TTATCGGCTTGTGCAACGG	XhoI
529-(His/GST)	Fwd	CGCGGATCCGCTAGC-TCCGGCAGCAAAACCGA	Bam HI-NheI
(MC58)	Rev	GCCCAAGCTT-ACGCAGTTCGGAATGGAG	HindIII
552L	Fwd	GCCGCCATATGTTGAATATTAAACTGAAAACCTTG	NdeI
	Rev	GCCGCCTCGAGTTATTTCTGATGCCTTTTCCC	XhoI
556L	Fwd	GCCGCCATATGGACAATAAGACCAAACTG	NdeI
	Rev	GCCGCCTCGAGTTAACGGTGCGGACGTTTC	XhoI
557L	Fwd	CGCGGATCCCATATG-AACAAACTGTTTCTTAC	NdeI
	Rev	CCCGCTCGAG-TCATTCCGCCTTCAGAAA	XhoI
564ab-(His/GST)	Fwd	CGCGGATCCCATATG-	BamHI-NdeI
(MC58)		CAAGGTATCGTTGCCGACAAATCCGCACCT	
	Rev	CCCGCTCGAG-	XhoI
ECA-LI (MCE9)	Fwd	AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	NdeI
564abL (MC58)	rwa	CGCGGATCC <u>CATATG</u> - AACCGCACCCTGTACAAAGTTGTATTTAACAAACATC	Nuci
	Rev	CCCGCTCGAG-	XhoI
		TTAAGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	
564b-	Fwd	CGCGGATCCCATATG- ACGGGAGAAAATCATGCGGTTTCACTTCATG	BamHI-NdeI
(His/GST)(MC58)	Rev	CCCGCTCGAG-	XhoI
	ROV	AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	241104
564c-	Fwd	CGC <u>GGATCCCATATG</u> -	BamHI-NdeI
(His/GST)(MC58)		GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	
	Rev	CCCG <u>CTCGAG</u> - GCGGTAACTGCCGCTTGCACTGAATCCGTAA	XhoI
564bc-	Fwd	CGCGGATCCCATATG-	BamHI-NdeI
(His/GST)(MC58)	1	ACGGAGAAAATCATGCGGTTTCACTTCATG	
	Rev	CCCG <u>CTCGAG</u> -	XhoI
		GCGGTAACTGCCGCTTGCACTGAATCCGTAA	D 1113117
564d- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- CAAAGCAAAGTCAAAGCAGACCATGCCTCCGTAA	BamHI-NdeI
(1115/ 001)(1/1000)	Rev	CCCGCTCGAG-	XhoI
		TCTTTCCTTTCAATTATAACTTTAGTAGGTTCAATTTTG	
		GTCCCC	
564cd- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
(III) (DI)(MICSO)	Rev	CCCGCTCGAG-	XhoI
	,	TCTTTCCTTTCAATTATAACTTTAGTAGGTTCAATTTTG	
		GTCCCC	
570L	Fwd	GCGGC <u>CATATG</u> -ACCCGTTTGACCCGCG	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAGCGGGCGTTCATTTCTT	XhoI
576-1L	Fwd	CGCGGATCC <u>CATATG</u> -AACACCATTTTCAAAATC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTAATTTACTTTTTTGATGTCG	XhoI

580L	Fwd	GCGGCCATATG-GATTCGCCCAAGGTCGG	NdeI
20017	Rev	GCGGC <u>CTCGAG</u> -CTACACTTCCCCCGAAGTGG	XhoI
583L	Fwd	CGCGGATCCCATATG-ATAGTTGACCAAAGCC	NdeI
30311	Rev	CCCGCTCGAG-TTATTTTTCCGATTTTTCGG	XhoI
593	Fwd	GCGGC <u>CATATG</u> -CTTGAACTGAACGGACT	NdeI
393	Rev	GCGGC <u>CTCGAG</u> -TCAGCGGAAGCGGACGATT	XhoI
650 (His-GST)	Fwd		
(MC58)	1·wu	CGC <u>GGATCCCATATG</u> TCCAAAACTCAAAACCATCG	BamHI-Ndel
, ,	Rev	CCCGCTCGAGGCTTCCAATCAGTTTGACC	XhoI
652	Fwd	GCGGC <u>CATATG</u> -AGCGCAATCGTTGATATTTTC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTATTTGCCCAGTTGGTAGAATG	XhoI
664L	Fwd	GCGGC <u>CATATG</u> -GTGATACATCCGCACTACTTC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAAAATCGAGTTTTACACCA	XhoI
726	Fwd	GCGGC <u>CATATG</u> -ACCATCTATTTCAAAAACGG	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAGCCGATGTTTAGCGTCCATT	XhoI
741-His(MC58)	Fwd	CGCGGATCC <u>CATATG</u> -AGCAGCGGAGGGGGTG	NdeI
·	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAGGC	XhoI
ΔG741-His(MC58)	Fwd	CGCGGATCC <u>CATATG-</u> GTCGCCGCCGACATCG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAGGC	XhoI
686-2-(His/GST)	Fwd	CGCGGATCCCATATG-GGCGGTTCGGAAGGCG	BamHI-Ndel
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTGAACACTGATGTCTTTTCCGA	XhoI
719-(His/GST)	Fwd	CGC <u>GGATCCGCTAGC</u> -AAACTGTCGTTGGTGTTAAC	BamHI-Nhel
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTGACCCGCTCCACGG	XhoI
730-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCCGCCTCGAGATCTCCTAAACCTGTTTTAACAATGCCG	XhoI
730A-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGCTCCATGCTGTTGCCCCAGC	XhoI
730B-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGAAAATCCCCGCTAACCGCAG	XhoI
741-His	Fwd	CGCGGATCC <u>CATATG</u> -AGCAGCGGAGGGGGTG	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAGGC	XhoI
ΔG741-His	Fwd	CGCGGATCC <u>CATATG</u> -GTCGCCGCCGACATCG	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAGGC	XhoI
743 (His-GST)	Fwd	CGC <u>GGATCCCATATG</u> GACGGTGTTGTGCCTGTT	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> CTTACGGATCAAATTGACG	XhoI
757 (His-GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> GGCAGCCAATCTGAAGAA	BamHI-Ndel
	Rev	CCCGCTCGAGCTCAGCTTTTGCCGTCAA	XhoI
759-His/GST	Fwd	CGCGGATCCGCTAGC-TACTCATCCATTGTCCGC	BamHI-Nhel
(MC58)	Rev	CCCG <u>CTCGAG</u> -CCAGTTGTAGCCTATTTTG	XhoI
759L	Fwd	CGCGGATCCGCTAGC-ATGCGCTTCACACACAC	NheI
(MC58)	Rev	CCCGCTCGAG-TTACCAGTTGTAGCCTATTT	XhoI
760-His	Fwd	GCCGCCATATGGCACAAACGGAAGGTTTGGAA	NdeI
	Rev	GCCGCCTCGAGAAAACTGTAACGCAGGTTTGCCGTC	XhoI
769-His (MC58)	Fwd	GCGGCCATATGGAAGAACACCGCGCGAACCG	NdeI

	Rev	GCGGCCTCGAGGAACGTTTTATTAAACTCGAC	XhoI
907L	Fwd	GCGGC <u>CATATG</u> -AGAAAACCGACCGATACCCTA	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TCAACGCCACTGCCAGCGGTTG	XhoI
911L	Fwd	CGCGGATCC <u>CATATG</u> -AAGAAGAACATATTGGAATTTTGGGTCGGACTG	NdeI
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LOmpA	Fwd	GGGAATTC <u>CATATG</u> AAAAAGACAGCTATCGCGATTGCA GTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCC <u>GC</u> <u>TAGC</u> -GCTTTCCGCGTGGCCGGCGGTGC	NdeI-(NheI)
	Rev	CCCG <u>CTCGAG</u> -TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LPelB	Fwd	CATG <u>CCATGG</u> -CTTTCCGCGTGGCCGGCGGTGC	NcoI
	Rev	CCCG <u>CTCGAG</u> -TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
913-His/GST	Fwd	CGC <u>GGATCCCATATG</u> -TTTGCCGAAACCCGCC	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -AGGTTGTGTTCCAGGTTG	XhoI
913L	Fwd	CGCGGATCC <u>CATATG</u> -AAAAAAACCGCCTATG	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTAAGGTTGTGTTCCAGG	XhoI
919L	Fwd	CGCGGATCC <u>CATATG</u> -AAAAAATACCTATTCCGC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACGGGCGGTATTCGG	XhoI
919	Fwd	CGCGGATCC <u>CATATG-</u> CAAAGCAAGAGCATCCAAA	NdeI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCGG	XhoI
919L Orf4	Fwd	GGGAATTC <u>CATATG</u> AAAACCTTCTTCAAAACCCTTTCCG CCGCCGC <u>GCTAGC</u> GCTCATCCTCGCCGCC- TGCCAAAGCAAGAGCATC	NdeI-(NheI)
	Rev	CCCG <u>CTCGAG</u> -TTACGGGCGGTATTCGGGCTTCATACCG	XhoI
(919)-287fusion	Fwd	CGCGGATCC <u>GTCGAC-</u> TGTGGGGGCGGCGGTGGC	SalI
	Rev	CCCG <u>CTCGAG</u> -TCAATCCTGCTCTTTTTTGCC	XhoI
920-1L	Fwd	GCGGC <u>CATATG</u> -AAGAAAACATTGACACTGC	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTAATGGTGCGAATGACCGAT	XhoI
925-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctTGCGGCAAGGATGCCGG	attB1
	Rev	ggggaccactttgtacaagaaagctgggtCTAAAGCAACAATGCCGG	attB2
926L	Fwd	CGCGGATCC <u>CATATG</u> -AAACACACCGTATCC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATCTCGTGCGCGCC	XhoI
927-2-(His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -AGCCCCGCGCCGATT	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTTTTGTGCGGTCAGGCG	XhoI
932-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctTGTTCGTTTGGGGGATTTAA ACCAAACCAAATC	attB1
935 (His-GST) (MC58)	For	CGC <u>GGATCCCATATG</u> GCGGATGCGCCCGCG	BamHI-NdeI
·	Rev	CCCGCTCGAGAAACCGCCAATCCGCC	XhoI
	Rev	ggggaccactttgtacaagaaagctgggtTCATTTTGTTTTTCCTTCTTCTCGAGGCCATT	attB2
936-1L	Fwd	CGCGGATCC <u>CATATG</u> -AAACCCAAACCGCAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAGCGTTGGACGTAGT	XhoI
953L	Fwd	GGGAATTC <u>CATATG</u> -AAAAAAATCATCTTCGCCG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTGTTTGGCTGCCTCGAT	XhoI
953-fu	Fwd	GGGAATTC <u>CATATG</u> -GCCACCTACAAAGTGGACG	NdeI
	Rev	CGG <u>GGATCC</u> -TTGTTTGGCTGCCTCGATTTG	BamHI

954 (His-GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> CAAGAACAATCGCAGAAAG	BamHI-NdeI
, .	Rev	CCCGCTCGAGTTTTTTCGGCAAATTGGCTT	XhoI
958-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctGCCGATGCCGTTGCGG	attB1
	Rev	ggggaccactttgtacaagaaagctgggtTCAGGGTCGTTTGTTGCG	attB2
961L	Fwd	CGCGGATCC <u>CATATG</u> -AAACACTTTCCATCC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAGCGACGAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961 c (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -GCCACAAACGACG	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -ACCCACGTTGTAAGGTTG	XhoI
961 c-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -ACCCACGTTGTAAGGTTG	XhoI
961 c-L	Fwd	CGCGGATCC <u>CATATG</u> -ATGAAACACTTTCCATCC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTAACCCACGTTGTAAGGT	XhoI
961 c-L	Fwd	CGCGGATCC <u>CATATG</u> -ATGAAACACTTTCCATCC	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTAACCCACGTTGTAAGGT	XhoI
961 d (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -GCCACAAACGACG	BamHI-NdeI
	Rev	CCCGCTCGAG-GTCTGACACTGTTTTATCC	XhoI
961 Δ1-L	Fwd	CGCGGATCC <u>CATATG</u> -ATGAAACACTTTCCATCC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATGCTTTGGCGGCAAAG	XhoI
fu 961	Fwd	CGCGGATCC <u>CATATG</u> - GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAGCGACGAC	NdeI
(MC58)	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961 c	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu 961 c-L	Fwd	CGCGGATCC <u>CATATG</u> - ATGAAACACTTTCCATCC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu (961)- 741(MC58)-His	Fwd	CGC <u>GGATCC</u> -GGAGGGGGTGTCG	BamHI
	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAGGC	XhoI
fu (961)-983-His	Fwd	CGC <u>GGATCC</u> - GGCGGAGGCGCACTT	BamHI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
fu (961)- Orf46.1- His	Fwd	CGC <u>GGATCC</u> GGTGGTGGT- TCAGATTTGGCAAACGATTC	BamHI
1115	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI
fu (961 c-L)-	Fwd	CGCGGATCC -GGAGGGGGTGTCG	BamHI
741(MC58)	Rev	CCCGCTCGAG-TTATTGCTTGGCGGCAAG	XhoI
fu (961c-L)-983	Fwd	CGCGGATCC - GGCGGAGGCGCACTT	BamHI
iu (2010-12)*703	Rev	CCCGCTCGAG-TCAGAACCGGTAGCCTAC	XhoI
fu (961c-L)- Orf46.1	Fwd	CGCGGATCCGGTGGTGGTGT- TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
961-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACG	BamHI-NdeI

(MC58)	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
961 Δ1-His	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACGAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TGCTTTGGCGGCAAAGTT	XhoI
961a-(His/GST)	Fwd	CGCGGATCC <u>CATATG</u> -GCCACAAACGACGAC	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -TTTAGCAATATTATCTTTGTTCGTAGC	XhoI
961b-(His/GST)	Fwd	CGCGGATCC <u>CATATG</u> -AAAGCAAACCGTGCCGA	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
961-His/GST GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctGCAGCCACAAACGACGACGATGTTAAAAAAGC	attB1
	Rev	ggggaccactttgtacaagaaagctgggtTTACCACTCGTAATTGACGCCGACATGGTAGG	attB2
982	Fwd	GCGGC <u>CATATG</u> -GCAGCAAAAGACGTACAGTT	NdeI
	Rev	GCGGC <u>CTCGAG</u> -TTACATCATGCCGCCCATACCA	XhoI
983-His (2996)	Fwd	CGCGGATCC <u>GCTAGC</u> -TTAGGCGGCGGCGGAG	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
ΔG983-His (2996)	Fwd	CCCCTAGCTAGC-ACTTCTGCGCCCGACTT	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
983-His	Fwd	CGCGGATCC <u>GCTAGC</u> -TTAGGCGGCGGCGAG	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
ΔG983-His	Fwd	CGCGGATCC <u>GCTAGC</u> -ACTTCTGCGCCCGACTT	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
983L	Fwd	CGCGGATCC <u>GCTAGC</u> - CGAACGACCCCAACCTTCCCTACAAAAACTTTCAA	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAGAACCGACGTGCCAAGCCGTTC	XhoI
987-His (MC58)	Fwd	GCCGCCATATGCCCCCACTGGAAGAACGGACG	NdeI
	Rev	GCCGCCTCGAGTAATAAACCTTCTATGGGCAGCAG	XhoI
989-(His/GST)	Fwd	CGCGGATCCCATATG-TCCGTCCACGCATCCG	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTTGAATTTGTAGGTGTATTG	XhoI
989L	Fwd	CGCGGATCC <u>CATATG</u> -ACCCCTTCCGCACT	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTATTTGAATTTGTAGGTGTAT	XhoI
CrgA-His	Fwd	CGCGGATCC <u>CATATG</u> -AAAACCAATTCAGAAGAA	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TCCACAGAGATTGTTTCC	XhoI
PilC1-ES	Fwd	GATGCCCGAAGGGCGGG	
(MC58)	Rev	GCCC <u>AAGCTT</u> -TCAGAAGAAGACTTCACGC	
PilC1-His	Fwd	CGCGGATCC <u>CATATG</u> -CAAACCCATAAATACGCTATT	NdeI
(MC58)	Rev	GCCC <u>AAGCTT</u> -GAAGAAGACTTCACGCCAG	HindIII
Δ1PilC1-His	Fwd	CGCGGATCC <u>CATATG</u> -GTCTTTTTCGACAATACCGA	NdeI
(MC58)	Rev	GCCC <u>AAGCTT</u> -	HindIII
PilC1L	Fwd	CGCGGATCC <u>CATATG</u> -AATAAAACTTTAAAAAGGCGG	NdeI
(MC58)	Rev	GCCC <u>AAGCTT</u> -TCAGAAGAAGACTTCACGC	HindIII
ΔGTbp2-His	Fwd	CGCGAATCC <u>CATATG</u> -TTCGATCTTGATTCTGTCGA	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His	Fwd	CGCGAATCC <u>CATATG</u> -TTGGGCGGAGGCGGCAG	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His(MC58)	Fwd	CGCGAATCC <u>CATATG</u> -TTGGGCGGAGGCGGCAG	NdeI
	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI

NMB0109- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -GCAAATTTGGAGGTGCGC	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTCGGAGCGGTTGAAGC	XhoI
NMB0109L	Fwd	CGCGGATCC <u>CATATG</u> -CAACGTCGTATTATAACCC	NdeI
(MC58)	Rev CCCGCTCGAG-TTATTCGGAGCGGTTGAAG	XhoI	
NMB0207- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> - GGCATCAAAGTCGCCATCAACGGCTAC	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTTGAGCGGGCGCACTTCAAGTCCG	XhoI
NMB0462- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -GGCGCGCGCGAAAAAAAC	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -GTTGGTGCCGACTTTGAT	XhoI
NMB0623- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -GGCGGCGGAAGCGATA	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TTTGCCCGCTTTGAGCC	XhoI
NMB0625 (His- GST)(MC58)	Fwd	CGC <u>GGATCCCATATG</u> GGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGCATCCCGTACTGTTTCG	XhoI
NMB0634 (His/GST)(MC58)	Fwd	ggggacaagtttgtacaaaaaagcaggctCCGACATTACCGTGTACAAC GGCCAACAAAGAA	attB1
	Rev	ggggaccactttgtacaagaaagctgggtCTTATTTCATACCGGCTTGCT CAAGCAGCCGG	attB2
NMB0776- His/GST (MC58)	Fwd	ggggacaagtttgtacaaaaaagcaggctGATACGGTGTTTTCCTGTAA AACGGACAACAA	attB1
OA EE	Rev	ggggaccactttgtacaagaaagctgggtCTAGGAAAAATCGTCATCGT TGAAATTCGCC	attB2
NMB1115-	Fwd	ggggacaagtttgtacaaaaaagcaggctATGCACCCATCGAAACC	attB1
His/GST (MC58) GATE	Rev	ggggaccactttgtacaagaaagctgggtCTAGTCTTGCAGTGCCTC	attB2
NMB1343- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> - GGAAATTTCTTATATAGAGGCATTAG	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> - GTTAATTTCTATCAACTCTTTAGCAATAAT	XhoI
NMB1369 (His- GST (MC58)	Fwd	CGC <u>GGATCCCATATG</u> GCCTGCCAAGACGACA	BamHI-NdeI
	Rev	CCCGCTCGAGCCGCCTCCTGCCGAAA	XhoI
NMB1551 (His- GST)(MC58)	Fwd	CGC <u>GGATCCCATATG</u> GCAGAGATCTGTTTGATAA	BamHI-NdeI
	Rev	CCCGCTCGAGCGGTTTTCCGCCCAATG	XhoI
NMB1899 (His- GST) (MC58)	Fwd	CGC <u>GGATCCCATATG</u> CAGCCGGATACGGTC	BamHI-NdeI
	Rev	CCCGCTCGAGAATCACTTCCAACACAAAAT	XhoI
NMB2050- (His/GST)	Fwd	CGC <u>GGATCCCATATG</u> -TGGTTGCTGATGAAGGGC	BamHI-NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -GACTGCTTCATCTTCTGC	XhoI
NMB2050L	Fwd	CGCGGATCC <u>CATATG</u> -GAACTGATGACTGTTTTGC	NdeI
(MC58)	Rev	CCCG <u>CTCGAG</u> -TCAGACTGCTTCATCTTCT	XhoI
NMB2159-	Fwd	CGCGGATCCCATATG-	BamHI-NdeI
(His/GST) (MC58)	Rev	AGCATTAAAGTAGCGATTAACGGTTTCGGC CCCGCTCGAG-	XhoI
(141030)	100	GATTTTGCCTGCGAAGTATTCCAAAGTGCG	ZXIIOI
fu-∆G287His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI

	Rev	CGG <u>GGATCC</u> -ATCCTGCTCTTTTTTGCCGG	BamHI
fu-(ΔG287)-919- His	Fwd	CGC <u>GGATCC</u> GGTGGTGGT- CAAAGCAAGAGCATCCAAACC	BamHI
	Rev	CCC <u>AAGCTT</u> -TTCGGGCGGTATTCGGGCTTC	HindIII
fu-(ΔG287)-953- His	Fwd	CGC <u>GGATCC</u> GGTGGTGGT- GCCACCTACAAAGTGGAC	BamHI
	Rev	GCCC <u>AAGCTT</u> -TTGTTTGGCTGCCTCGAT	HindIII
fu-(ΔG287)-961-	Fwd	CGC <u>GGATCC</u> GGTGGTGGTGGT-ACAAGCGACGACG	BamHI
His	Rev	GCCC <u>AAGCTT</u> -CCACTCGTAATTGACGCC	HindIII
fu-(ΔG287)- Orf46.1-His	Fwd	CGC <u>GGATCC</u> GGTGGTGGT- TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCC <u>AAGCTT</u> -CGTATCATATTTCACGTGC	HindIII
fu-(ΔG287-919)- Orf46.1-His	Fwd	CCC <u>AAGCTT</u> GGTGGTGGTGGT- TCAGATTTGGCAAACGATTC	HindIII
	Rev	CCC <u>GCTCGAG</u> -CGTATCATATTTCACGTGC	XhoI
fu-(ΔG287- Orf46.1)-919-His	Fwd	CCC <u>AAGCTT</u> GGTGGTGGTGGT- CAAAGCAAGAGCATCCAAACC	HindIII
·	Rev	CCC <u>GCTCGAG</u> -CGGGCGGTATTCGGGCTT	XhoI
fu ΔG287(394.98)- 	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC	NheI
	Rev	CGG <u>GGATCC</u> -ATCCTGCTCTTTTTTGCCGG	BamHI
fu Orf1-(Orf46.1)-	Fwd	CGCGGATCC <u>GCTAGC</u> -GGACACACTTATTTCGGCATC	NheI
His	Rev	CGCGGATCC-CCAGCGGTAGCCTAATTTGAT	
fu (Orf1)-Orf46.1- His	Fwd	CGC <u>GGATCC</u> GGTGGTGGT- TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCC <u>AAGCTT</u> -CGTATCATATTTCACGTGC	HindIII
fu (919)-Orf46.1-	Fwd1	GCGGC <u>GTCGAC</u> GGTGGCGGAGGCACTGGATCCTCAG	SalI
His	Fwd2	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC	
	Rev	CCC <u>GCTCGAG</u> -CGTATCATATTTCACGTGC	XhoI
Fu orf46	Fwd	GGAATTC <u>CATATG</u> TCAGATTTGGCAAACGATTC	NdeI
	Rev	CGC <u>GGATCC</u> CGTATCATATTTCACGTGC	BamHI
Fu (orf46)-287-His	Fwd	CGG <u>GGATCC</u> GGGGGCGCGTGGCG	BamHI
	Rev	CCC <u>AAGCTT</u> ATCCTGCTCTTTTTTGCCGGC	HindIII
Fu (orf46)-919-His	Fwd	CGC <u>GGATCC</u> GGTGGTGGTGAAAGCAAGAGCATCCA AACC	
	Rev	CCC <u>AAGCTT</u> CGGGCGTATTCGGGCTTC	HindIII
Fu (orf46-919)- 287-His	Fwd	CCCC <u>AAGCTT</u> GGGGGCGGCGGTGGCG	HindIII
	Rev	CCCGCTCGAGATCCTGCTCTTTTTTGCCGGC	XhoI
Fu (orf46-287)- 919-His	Fwd	CCC <u>AAGCTT</u> GGTGGTGGTGGTCAAAGCAAGAGCAT CCAAACC	HindIII
	Rev	CCCGCTCGAGCGGCGTATTCGGGCTT	XhoI
		CCA CCCA CTCCA ATTCCCCCA CCCA CCA ACCA CCA	XhoI
(ΔG741)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	1202
(ΔG741)-961c-His	Fwd2	GCGGC <u>CTCGAG</u> -GGTGGCGGAGGCACTGGATCCGCAG	
	Fwd2 Rev	GCGGC <u>CTCGAG</u> -GGTGGCGGAGGCACTGGATCCGCAG CCCG <u>CTCGAG</u> -ACCCAGCTTGTAAGGTTG	XhoI
(ΔG741)-961c-His (ΔG741)-961-His	Fwd2	GCGGC <u>CTCGAG</u> -GGTGGCGGAGGCACTGGATCCGCAG CCCG <u>CTCGAG</u> -ACCCAGCTTGTAAGGTTG GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	

-92-

(ΔG741)-983-His	Fwd	GCGGC <u>CTCGAG</u> - GGATCCGGCGGAGGCGCACTTCTGCG	XhoI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
(ΔG741)-orf46.1- His	Fwd1 Fwd2	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC GCGGC <u>GTCGAC</u> GGTGGCGGAGGCACTGGATCCTCAGA	SalI
	Rev	CCCG <u>CTCGAG</u> -CGTATCATATTTCACGTGC	XhoI
(ΔG983)- 741(MC58) -His	Fwd	GCGGC <u>CTCGAG</u> -GGATCCGGAGGGGGTGGTGTCGCC	XhoI
	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAG	XhoI
(ΔG983)-961c-His	Fwd1 Fwd2	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA GCGGC <u>CTCGAG</u> -GGTGGCGGAGGCACTGGATCCGCAG	XhoI
	Rev	CCCG <u>CTCGAG</u> -ACCCAGCTTGTAAGGTTG	XhoI
(ΔG983)-961-His	Fwd1 Fwd2	GGAGGCACTGGATCCGCAGCCACAAACGACGA GCGGC <u>CTCGAG</u> -GGTGGCGGAGGCACTGGATCCGCAG	XhoI
	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
(ΔG983)- Orf46.1- His	Fwd1 Fwd2	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC GCGGC <u>GTCGAC</u> GGTGGCGGAGGCACTGGATCCTCAGA	SalI
	Rev	CCCG <u>CTCGAG</u> -CGTATCATATTTCACGTGC	XhoI

^{*} This primer was used as a Reverse primer for all the C terminal fusions of 287 to the His-tag.

§ Forward primers used in combination with the 287-His Reverse primer.

NB – All PCR reactions use strain 2996 unless otherwise specified (e.g. strain MC58)

In all constructs starting with an ATG not followed by a unique *NheI* site, the ATG codon is part of the *NdeI* site used for cloning. The constructs made using *NheI* as a cloning site at the 5' end (*e.g.* all those containing 287 at the N-terminus) have two additional codons (GCT AGC) fused to the coding sequence of the antigen.

Preparation of chromosomal DNA templates

5

10

15

N.meningitidis strains 2996, MC58, 394.98, 1000 and BZ232 (and others) were grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% w/v sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml of lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50μg/ml Proteinase K), and the suspension incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one CHCl₃/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes of ethanol, and collected by centrifugation. The pellet was washed once with 70%(v/v) ethanol and redissolved in 4.0ml TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). The DNA concentration was measured by reading OD₂₆₀.

PCR Amplification

5

10

15

20

30

The standard PCR protocol was as follows: 200ng of genomic DNA from 2996, MC581000, or BZ232 strains or 10ng of plasmid DNA preparation of recombinant clones were used as template in the presence of 40µM of each oligonucletide primer, 400-800 µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ, Boerhingher Mannheim ExpandTM Long Template).

After a preliminary 3 minute incubation of the whole mix at 95°C, each sample underwent a two-step amplification: the first 5 cycles were performed using the hybridisation temperature that excluded the restriction enzyme tail of the primer (T_{m1}) . This was followed by 30 cycles according to the hybridisation temperature calculated for the whole length oligos (T_{m2}) . Elongation times, performed at 68°C or 72°C, varied according to the length of the Orf to be amplified. In the case of Orf1 the elongation time, starting from 3 minutes, was increased by 15 seconds each cycle. The cycles were completed with a 10 minute extension step at 72°C.

The amplified DNA was either loaded directly on a 1% agarose gel. The DNA fragment corresponding to the band of correct size was purified from the gel using the Qiagen Gel Extraction Kit, following the manufacturer's protocol.

Digestion of PCR fragments and of the cloning vectors

The purified DNA corresponding to the amplified fragment was digested with the appropriate restriction enzymes for cloning into pET-21b+, pET22b+ or pET-24b+. Digested fragments were purified using the QIAquick PCR purification kit (following the manufacturer's instructions) and eluted with either H₂O or 10mM Tris, pH 8.5. Plasmid vectors were digested with the appropriate restriction enzymes, loaded onto a 1.0% agarose gel and the band corresponding to the digested vector purified using the Qiagen QIAquick Gel Extraction Kit.

25 Cloning

The fragments corresponding to each gene, previously digested and purified, were ligated into pET21b+, pET22b+ or pET-24b+. A molar ratio of 3:1 fragment/vector was used with T4 DNA ligase in the ligation buffer supplied by the manufacturer.

Recombinant plasmid was transformed into competent *E.coli* DH5 or HB101 by incubating the ligase reaction solution and bacteria for 40 minutes on ice, then at 37°C for 3 minutes.

This was followed by the addition of 800µl LB broth and incubation at 37°C for 20 minutes. The cells were centrifuged at maximum speed in an Eppendorf microfuge, resuspended in approximately 200µl of the supernatant and plated onto LB ampicillin (100mg/ml) agar.

Screening for recombinant clones was performed by growing randomly selected colonies overnight at 37°C in 4.0ml of LB broth + 100µg/ml ampicillin. Cells were pelleted and plasmid DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions. Approximately 1µg of each individual miniprep was digested with the appropriate restriction enzymes and the digest loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1kb DNA Ladder, GIBCO). Positive clones were selected on the basis of the size of insert.

Expression

5

10

15

20

25

30

After cloning each gene into the expression vector, recombinant plasmids were transformed into E.coli strains suitable for expression of the recombinant protein. 1µl of each construct was used to transform E.coli BL21-DE3 as described above. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, to give an OD₆₀₀ between 0.1 and 0.2. The flasks were incubated at 30°C or at 37°C in a gyratory water bath shaker until OD₆₀₀ indicated exponential growth suitable for induction of expression (0.4-0.8 OD). Protein expression was induced by addition of 1.0mM IPTG. After 3 hours incubation at 30°C or 37°C the OD₆₀₀ was measured and expression examined. 1.0ml of each sample was centrifuged in a microfuge, the pellet resuspended in PBS and analysed by SDS-PAGE and Coomassie Blue staining.

Gateway cloning and expression

Sequences labelled GATE were cloned and expressed using the GATEWAY Cloning Technology (GIBCO-BRL). Recombinational cloning (RC) is based on the recombination reactions that mediate the integration and excision of phage into and from the *E.coli* genome, respectively. The integration involves recombination of the *attP* site of the phage DNA within the *attB* site located in the bacterial genome (BP reaction) and generates an integrated phage genome flanked by *attL* and *attR* sites. The excision recombines *attL* and *attR* sites back to *attP* and *attB* sites (LR reaction). The integration reaction requires two enzymes [the phage protein Integrase (Int) and the bacterial protein integration host factor (IHF)] (BP clonase). The

excision reaction requires Int, IHF, and an additional phage enzyme, Excisionase (Xis) (LR clonase). Artificial derivatives of the 25-bp bacterial *attB* recombination site, referred to as B1 and B2, were added to the 5' end of the primers used in PCR reactions to amplify Neisserial ORFs. The resulting products were BP cloned into a "Donor vector" containing complementary derivatives of the phage *attP* recombination site (P1 and P2) using BP clonase. The resulting "Entry clones" contain ORFs flanked by derivatives of the *attL* site (L1 and L2) and were subcloned into expression "destination vectors" which contain derivatives of the *attL*-compatible *attR* sites (R1 and R2) using LR clonase. This resulted in "expression clones" in which ORFs are flanked by B1 and B2 and fused in frame to the GST or His N terminal tags.

The *E. coli* strain used for GATEWAY expression is BL21-SI. Cells of this strain are induced for expression of the T7 RNA polymerase by growth in medium containing salt (0.3 M NaCl).

Note that this system gives N-terminus His tags.

Preparation of membrane proteins.

5

15

20

25

30

Fractions composed principally of either inner, outer or total membrane were isolated in order to obtain recombinant proteins expressed with membrane-localisation leader sequences. The method for preparation of membrane fractions, enriched for recombinant proteins, was adapted from Filip *et. al.* [*J.Bact.* (1973) 115:717-722] and Davies *et. al.* [*J.Immunol.Meth.* (1990) 143:215-225]. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD₅₅₀ reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C and resuspended in 20 ml of 20 mM Tris-HCl (pH 7.5) and complete protease inhibitors (Boehringer-Mannheim). All subsequent procedures were performed at 4°C or on ice.

Cells were disrupted by sonication using a Branson Sonifier 450 and centrifuged at 5000g for 20 min to sediment unbroken cells and inclusion bodies. The supernatant, containing membranes and cellular debris, was centrifuged at 50000g (Beckman Ti50, 29000rpm) for 75 min, washed with 20 mM Bis-tris propane (pH 6.5), 1.0 M NaCl, 10% (v/v) glycerol and sedimented again at 50000g for 75 minutes. The pellet was resuspended in 20mM Tris-HCl (pH 7.5), 2.0% (v/v) Sarkosyl, complete protease inhibitor (1.0 mM EDTA, final

-96-

concentration) and incubated for 20 minutes to dissolve inner membrane. Cellular debris was pelleted by centrifugation at 5000g for 10 min and the supernatant centrifuged at 75000g for 75 minutes (Beckman Ti50, 33000rpm). Proteins 008L and 519L were found in the supernatant suggesting inner membrane localisation. For these proteins both inner and total membrane fractions (washed with NaCl as above) were used to immunise mice. Outer membrane vesicles obtained from the 75000g pellet were washed with 20 mM Tris-HCl (pH 7.5) and centrifuged at 75000g for 75 minutes or overnight. The OMV was finally resuspended in 500 µl of 20 mM Tris-HCl (pH 7.5), 10% v/v glycerol. Orf1L and Orf40L were both localised and enriched in the outer membrane fraction which was used to immunise mice. Protein concentration was estimated by standard Bradford Assay (Bio-Rad), while protein concentration of inner membrane fraction was determined with the DC protein assay (Bio-Rad). Various fractions from the isolation procedure were assayed by SDS-PAGE.

Purification of His-tagged proteins

5

10

15

20

25

30

Various forms of 287 were cloned from strains 2996 and MC58. They were constructed with a C-terminus His-tagged fusion and included a mature form (aa 18-427), constructs with deletions ($\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4$) and clones composed of either B or C domains. For each clone purified as a His-fusion, a single colony was streaked and grown overnight at 37°C on a LB/Amp (100 μg/ml) agar plate. An isolated colony from this plate was inoculated into 20ml of LB/Amp (100 µg/ml) liquid medium and grown overnight at 37°C with shaking. The overnight culture was diluted 1:30 into 1.0 L LB/Amp (100 µg/ml) liquid medium and allowed to grow at the optimal temperature (30 or 37°C) until the OD₅₅₀ reached 0.6-0.8. Expression of recombinant protein was induced by addition of IPTG (final concentration 1.0mM) and the culture incubated for a further 3 hours. Bacteria were harvested by centrifugation at 8000g for 15 min at 4°C. The bacterial pellet was resuspended in 7.5 ml of either (i) cold buffer A (300 mM NaCl, 50 mM phosphate buffer, 10 mM imidazole, pH 8.0) for soluble proteins or (ii) buffer B (10mM Tris-HCl, 100 mM phosphate buffer, pH 8.8 and, optionally, 8M urea) for insoluble proteins. Proteins purified in a soluble form included 287-His, $\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4287$ -His, $\Delta 4287$ MC58-His, 287c-His and 287cMC58-His. Protein 287bMC58-His was insoluble and purified accordingly. Cells were disrupted by sonication on ice four times for 30 sec at 40W using a Branson sonifier 450 and centrifuged at 13000xg for 30 min at 4°C. For insoluble proteins, pellets were resuspended in 2.0 ml buffer C (6 M guanidine hydrochloride, 100 mM phosphate buffer, 10 mM Tris- HCl, pH 7.5

and treated with 10 passes of a Dounce homogenizer. The homogenate was centrifuged at 13000g for 30 min and the supernatant retained. Supernatants for both soluble and insoluble preparations were mixed with 150µl Ni²⁺-resin (previously equilibrated with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 min. The resin was Chelating Sepharose Fast Flow (Pharmacia), prepared according to the manufacturer's protocol. The batch-wise preparation was centrifuged at 700g for 5 min at 4°C and the supernatant discarded. The resin was washed twice (batch-wise) with 10ml buffer A or B for 10 min, resuspended in 1.0 ml buffer A or B and loaded onto a disposable column. The resin continued to be washed with either (i) buffer A at 4°C or (ii) buffer B at room temperature, until the OD₂₈₀ of the flow-through reached 0.02-0.01. The resin was further washed with either (i) cold buffer C (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8.0) or (ii) buffer D (10mM Tris-HCl, 100mM phosphate buffer, pH 6.3 and, optionally, 8M urea) until OD₂₈₀ of the flow-through reached 0.02-0.01. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300 mM NaCl, 50mM phosphate buffer, 250 mM imidazole, pH 8.0) or (ii) elution buffer B (10 mM Tris-HCl, 100 mM phosphate buffer, pH 4.5 and, optionally, 8M urea) and fractions collected until the OD₂₈₀ indicated all the recombinant protein was obtained. 20µl aliquots of each elution fraction were analysed by SDS-PAGE. Protein concentrations were estimated using the Bradford assay.

20 Renaturation of denatured His-fusion proteins.

5

10

15

25

Denaturation was required to solubilize 287bMC8, so a renaturation step was employed prior to immunisation. Glycerol was added to the denatured fractions obtained above to give a final concentration of 10% v/v. The proteins were diluted to 200 μg/ml using dialysis buffer I (10% v/v glycerol, 0.5M arginine, 50 mM phosphate buffer, 5.0 mM reduced glutathione, 0.5 mM oxidised glutathione, 2.0M urea, pH 8.8) and dialysed against the same buffer for 12-14 hours at 4°C. Further dialysis was performed with buffer II (10% v/v glycerol, 0.5M arginine, 50mM phosphate buffer, 5.0mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was estimated using the formula:

Protein
$$(mg/ml) = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

PCT/IB01/00452

Amino acid sequence analysis.

Automated sequence analysis of the NH₂-terminus of proteins was performed on a Beckman sequencer (LF 3000) equipped with an on-line phenylthiohydantoin-amino acid analyser (System Gold) according to the manufacturer's recommendations.

-98-

5 Immunization

10

15

20

25

WO 01/64922

Balb/C mice were immunized with antigens on days 0, 21 and 35 and sera analyzed at day 49.

Sera analysis - ELISA

The acapsulated MenB M7 and the capsulated strains were plated on chocolate agar plates and incubated overnight at 37°C with 5% CO₂. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620} . The bacteria were let to grow until the OD reached the value of 0.4-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and bacteria were washed twice with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 1 hour at 37°C and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O₂) were added to each well and the plates were left at room temperature for 20 minutes. 100µl 12.5% H₂SO₄ was added to each well and OD490 was followed. The ELISA titers were calculated abitrarely as the dilution of sera which gave an OD₄₉₀ value of 0.4 above the level of preimmune sera. The ELISA was considered positive when the dilution of sera with OD₄₉₀ of 0.4 was higher than 1:400.

30 Sera analysis – FACS Scan bacteria binding assay

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C with 5% CO₂. Bacterial colonies were collected from the agar plates using

a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620} . The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA in PBS, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.05. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:100, 1:200, 1:400) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan (Laser Power 15mW) setting were: FL2 on; FSC-H threshold:92; FSC PMT Voltage: E 01; SSC PMT: 474; Amp. Gains 6.1; FL-2 PMT: 586; compensation values: 0.

Sera analysis – bactericidal assay

10

15

20

25

30

N. meningitidis strain 2996 was grown overnight at 37°C on chocolate agar plates (starting from a frozen stock) with 5% CO₂. Colonies were collected and used to inoculate 7ml Mueller-Hinton broth, containing 0.25% glucose to reach an OD_{620} of 0.05-0.08. The culture was incubated for approximately 1.5 hours at 37 degrees with shacking until the OD_{620} reached the value of 0.23-0.24. Bacteria were diluted in 50mM Phosphate buffer pH 7.2 containing 10mM MgCl₂, 10mM CaCl₂ and 0.5% (w/v) BSA (assay buffer) at the working dilution of 10^5 CFU/ml. The total volume of the final reaction mixture was 50 μ l with 25 μ l of serial two fold dilution of test serum, 12.5 μ l of bacteria at the working dilution, 12.5 μ l of baby rabbit complement (final concentration 25%).

Controls included bacteria incubated with complement serum, immune sera incubated with bacteria and with complement inactivated by heating at 56°C for 30'. Immediately after the addition of the baby rabbit complement, 10µl of the controls were plated on Mueller-Hinton agar plates using the tilt method (time 0). The 96-wells plate was incubated for 1 hour at 37°C with rotation. 7µl of each sample were plated on Mueller-Hinton agar plates as spots, whereas 10µl of the controls were plated on Mueller-Hinton agar plates using the tilt method

(time 1). Agar plates were incubated for 18 hours at 37 degrees and the colonies corresponding to time 0 and time 1 were counted.

PCT/IB01/00452

Sera analysis – western blots

WO 01/64922

5

10

30

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded onto a 12% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, using transfer buffer (0.3% Tris base, 1.44% glycine, 20% (v/v) methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

15 The OMVs were prepared as follows: N. meningitidis strain 2996 was grown overnight at 37 degrees with 5% CO₂ on 5 GC plates, harvested with a loop and resuspended in 10 ml of 20mM Tris-HCl pH 7.5, 2 mM EDTA. Heat inactivation was performed at 56°C for 45 minutes and the bacteria disrupted by sonication for 5 minutes on ice (50% duty cycle, 50%) output, Branson sonifier 3 mm microtip). Unbroken cells were removed by centrifugation at 20 5000g for 10 minutes, the supernatant containing the total cell envelope fraction recovered and further centrifuged overnight at 50000g at the temperature of 4°C. The pellet containing the membranes was resuspended in 2% sarkosyl, 20mM Tris-HCl pH 7.5, 2 mM EDTA and incubated at room temperature for 20 minutes to solubilise the inner membranes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, the supernatant 25 was further centrifuged at 50000g for 3 hours. The pellet, containing the outer membranes was washed in PBS and resuspended in the same buffer. Protein concentration was measured by the D.C. Bio-Rad Protein assay (Modified Lowry method), using BSA as a standard.

Total cell extracts were prepared as follows: *N. meningitidis* strain 2996 was grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

961 domain studies

<u>Cellular fractions preparation</u> Total lysate, periplasm, supernatant and OMV of *E.coli* clones expressing different domains of 961 were prepared using bacteria from over-night cultures or

-101-

after 3 hours induction with IPTG. Briefly, the periplasm were obtained suspending bacteria in saccarose 25% and Tris 50mM (pH 8) with polimixine 100μg/ml. After 1hr at room temperature bacteria were centrifuged at 13000rpm for 15 min and the supernatant were collected. The culture supernatant were filtered with 0.2μm and precipitated with TCA 50% in ice for two hours. After centrifugation (30 min at 13000 rp) pellets were rinsed twice with ethanol 70% and suspended in PBS. The OMV preparation was performed as previously described. Each cellular fraction were analyzed in SDS-PAGE or in Western Blot using the polyclonal anti-serum raised against GST-961.

5

10

Adhesion assay Chang epithelial cells (Wong-Kilbourne derivative, clone 1-5c-4, human conjunctiva) were maintained in DMEM (Gibco) supplemented with 10% heat-inactivated FCS, 15mM L-glutamine and antibiotics.

For the adherence assay, sub-confluent culture of Chang epithelial cells were rinsed with PBS and treated with trypsin-EDTA (Gibco), to release them from the plastic support. The cells were then suspended in PBS, counted and dilute in PBS to 5×10^5 cells/ml.

Bacteria from over-night cultures or after induction with IPTG, were pelleted and washed twice with PBS by centrifuging at 13000 for 5 min. Approximately 2-3x10⁸ (cfu) were incubated with 0.5 mg/ml FITC (Sigma) in 1ml buffer containing 50mM NaHCO₃ and 100mM NaCl pH 8, for 30 min at room temperature in the dark. FITC-labeled bacteria were wash 2-3 times and suspended in PBS at 1-1.5x10⁹/ml. 200µl of this suspension (2-3x10⁸)
were incubated with 200µl (1x10⁵) epithelial cells for 30min a 37°C. Cells were than centrifuged at 2000rpm for 5 min to remove non-adherent bacteria, suspended in 200µl of PBS, transferred to FACScan tubes and read

CLAIMS

- 1. A method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.
- 2. The method of claim 1, in which the protein of the invention is ORF46.
- 5 3. The method of claim 2, in which ORF46 is divided into a first domain (amino acids 1-433) and a second domain (amino acids 433-608).
 - 4. The method of claim 2, in which the protein of the invention is 564.
 - 5. The method of claim 4, in which protein 564 is divided into domains as shown in Figure 8.
- 10 6. The method of claim 1 in which the protein of the invention is 961.
 - 7. The method of claim 6, in which protein 961 is divided into domains as shown in Figure 12.
 - 8. The method of claim 1, in which the protein of the invention is 502 and the domain is amino acids 28 to 167 (numbered according to the MC58 sequence).
- 15 9. The method of claim 1, in which the protein of the invention is 287.
 - 10. A method for the heterologous expression of a protein of the invention, in which (a) a portion of the N-terminal domain of the protein is deleted.
 - 11. The method of claim 9 or claim 10, in which protein 287 is divided into domains A B & C shown in Figure 5.
- 20 12. The method of claim 11, in which (i) domain A, (ii) domains A and B, or (iii) domains A and C are deleted.
 - 13. The method of claim 11, wherein (i) amino acids 1-17, (ii) amino acids 1-25, (iii) amino acids 1-69, or (iv) amino acids 1-106, of domain A are deleted.
- 14. A method for the heterologous expression of a protein of the invention, in which (a) no fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

5

10

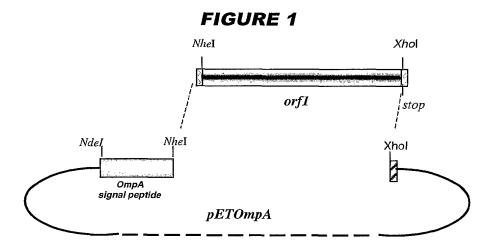
WO 01/64922 PCT/IB01/00452

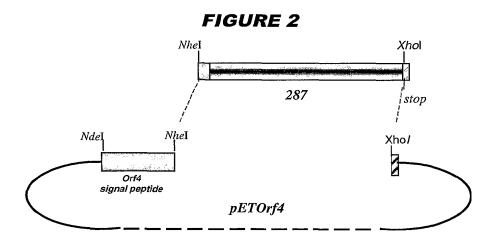
- 15. The method of claim 14, in which the protein of the invention is selected from the group consisting of: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109, NMB2050, 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.
- 16. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a different protein and, optionally, (b) no fusion partner is used.
- 17. The method of claim 16, in which the different protein is 961, ORF4, *E.coli* OmpA, or *E.carotovora* PelB, or in which the leader peptide is MKKYLFSAA.
- 18. The method of claim 17, in which the different protein is *E.coli* OmpA and the protein of the invention is ORF1.
- 15 19. The method of claim 17, in which the protein of the invention is 911 and the different protein is *E.carotovora* PelB or *E.coli* OmpA.
 - 20. The method of claim 17, in which the different protein is ORF4 and the protein of the invention is 287.
- 21. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used.
 - 22. The method of claim 21, in which the protein of the invention is 919.
 - 23. A method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.
- 25 24. The method of claim 23, in which protein 919 is expressed at 30°C.
 - 25. A method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

- WO 01/64922 PCT/IB01/00452
- 26. The method of claim 25, in which the protein of the invention is 907, 919 or 922.
- 27. The method of claim 26, in which 907 is mutated at Glu-117 (e.g. Glu→Gly).
- 28. The method of claim 26, in which 919 is mutated at Glu-255 (e.g. Glu→Gly) and/or Glu-323 (e.g. Glu→Gly).
- 5 29. The method of claim 26, in which 922 is mutated at Glu-164 (e.g. Glu→Gly), Ser-213 (e.g. Ser→Gly) and/or Asn-348 (e.g. Asn→Gly).
 - 30. A method for the heterologous expression of a protein of the invention, in which vector pSM214 is used or vector pET-24b is used.
- 31. The method of claim 30, in which the protein of the invention is 953 and the vector is pSM214.
 - 32. A method for the heterologous expression of a protein of the invention, in which a protein is expressed or purified such that it adopts a particular multimeric form.
 - 33. The method of claim 32, in which protein 953 is expressed and/or purified in monomeric form.
- 15 34. The method of claim 32, in which protein 961 is expressed and/or purified in tetrameric form.
 - 35. The method of claim 32, in which protein 287 is expressed and/or purified in dimeric form.
- 36. The method of claim 32, in which protein 919 is expressed and/or purified in monomeric form.
 - 37. A method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.
 - 38. The method of claim 37, in which the protein of the invention is 919, 287, ORF4, 406, 576, or ORF25.
- 39. A method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

-105-

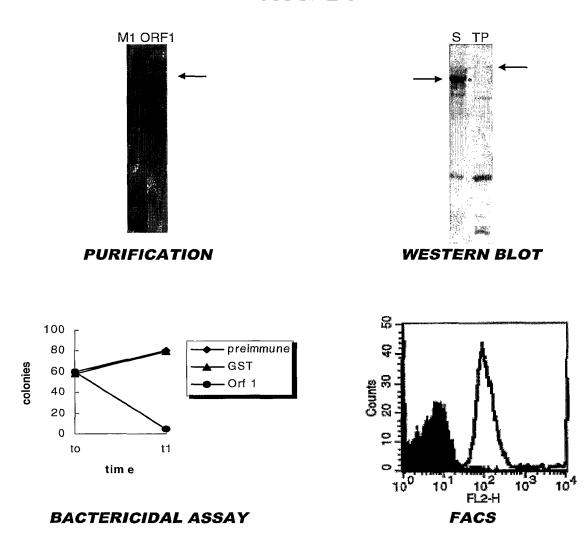
- 40. The method of claim 39, wherein the mutation is a substitution, an insertion, or a deletion
- 41. The method of claim 40, wherein the protein of the invention is 730, ORF29 or ORF46.
- 42. A method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.
- 5 43. The method of claim 42, in which the protein of the invention is 919.
 - 44. A method for the heterologous expression of a protein, in which a poly-glycine stretch within the protein is mutated.
 - 45. The method of claim 44, wherein the protein is a protein of the invention.
 - 46. The method of claim 45, wherein the protein of the invention is 287, 741, 983 or Tbp2.
- 10 47. The method of claim 46, wherein (Gly)₆ is deleted from 287 or 983.
 - 48. The method of claim 46, wherein (Gly)₄ is deleted from Tbp2 or 741
 - 49. The method of claim 47 or claim 48, wherein the leader peptide is also deleted.
 - 50. The method of any preceding claim, in which the heterologous expression is in an *E.coli* host.
- 15 51. A protein expressed by the method of any preceding claim.
 - 52. A heterologous protein comprising the N-terminal amino acid sequence MKKYLFSAA.





2/13

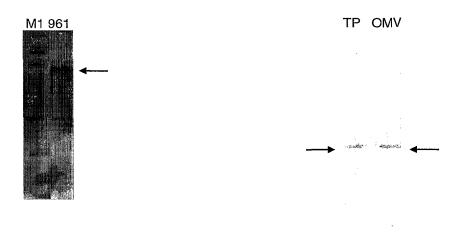
FIGURE 3



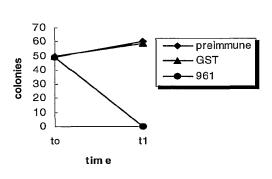
ELISA: POSITIVE

3/13

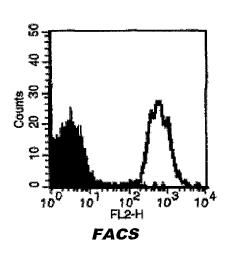
FIGURE 4



PURIFICATION



BACTERICIDAL ASSAY

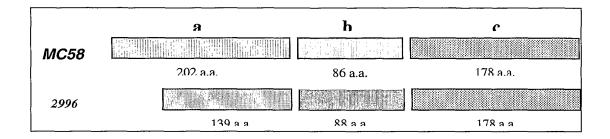


WESTERN BLOT

ELISA: POSITIVE

4/13

FIGURE 5



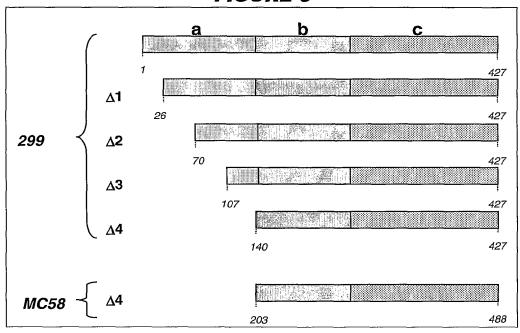
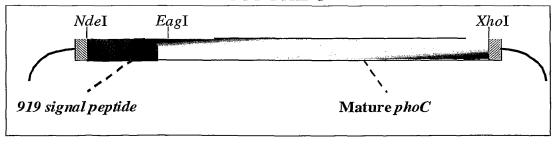


FIGURE 9

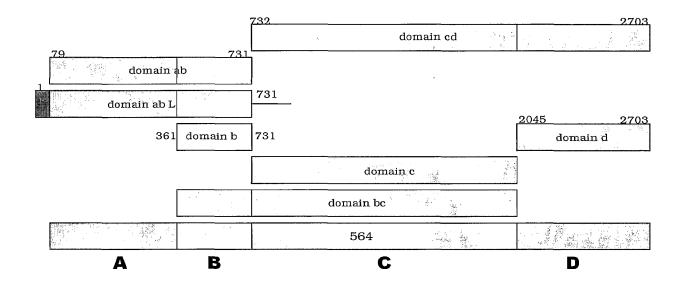


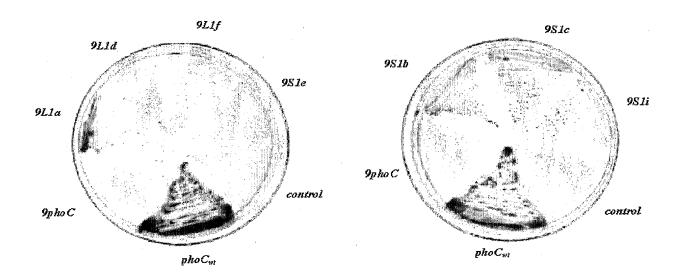
5/13

		<a< th=""></a<>
MC58 2996		mf <mark>k</mark> rsviamacifalsacggggggspdvksadtlskpaapvv <mark>s</mark> ekete <mark>a</mark> kedapqagsqg mf <mark>e</mark> rsviamacifalsacggggggspdvksadtlskpaapvv <mark>a</mark> ekete <mark>v</mark> kedapqagsqg
MC58	61	QGAPS <mark>AQGSQDMAAVS</mark> ENTGNGGA <mark>VTADN</mark> PKNEDEVAQNDMPQNAAGTDSSTPNHTPDP
2996	61	QGAPS <mark>T</mark> QGSQDMAAVS <mark>AENTGNGGAA</mark> TTDKPKNEDEGPQNDMPQN
MC58 2996	121 106	
MC58	181	AAGSSDPIPASNPAPANGGSNEGRVDLANGVLIDGPSQNITLTHCKGDSCSGNNELDEEV
2996	118	PADSSDSAPASNPAPANGGSNEGRVDLANGVLIDGPSQNITLTHCKGDSCNGDNLLDEEA
MC58	241	QIKSEFEKISDADKISNYKKDGKNDKEVGIVADSVQMKGINQYIIFYKPKPTSFARFR
2996	178	PSKSEFENINESERIEKYKKDGKSDKETNIVATAVQANGTNKYVIIYKDKSASSSSARFR
MC58 2996	299 238	<c< th=""></c<>
MC58	359	SYALRVQGEPAKGEMLAG <mark>A</mark> AVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDS
2996	298	SYALRVQGEPAKGEMLAG <mark>T</mark> AVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDS
MC58	419	GDDLHMGTQKFKAAIDGNGFKGTWTENG <mark>S</mark> GDVSG <mark>KFYGPAGEEVAGKYSYRPTDAEKGGF</mark>
2996	358	GDDLHMGTQKFKAAIDGNGFKGTWTENG <mark>G</mark> GDVSG <mark>R</mark> FYGPAGEEVAGKYSYRPTDAEKGGF
MC58 2996	479 418	C> GVFAGKKEQD* GVFAGKKEQD*

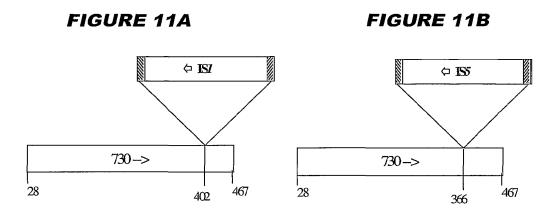
6/13

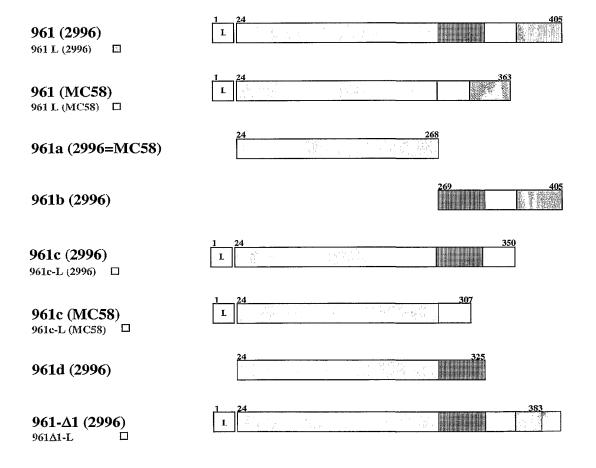
FIGURE 8





7/13





8/13

FIGURE 13

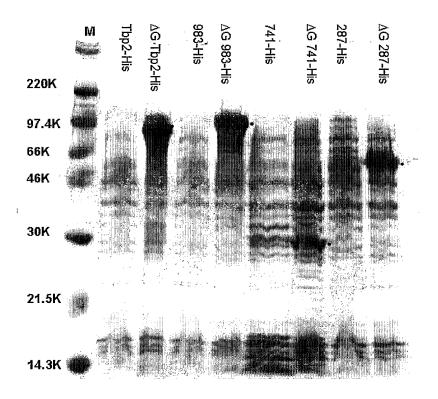


FIGURE 14

FIGURE 14A — ΔG287—919

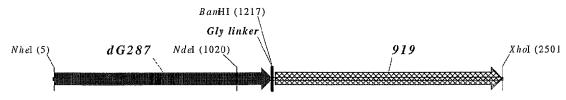


FIGURE 14B — ΔG287—953



9/13

FIGURE 14C — ΔG287—961

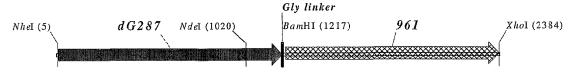


FIGURE 14D — ΔG287NZ—919

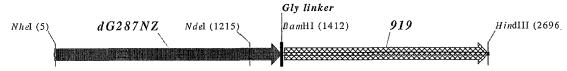


FIGURE 14E — ΔG287NZ—953



FIGURE 14F — ΔG287NZ—961

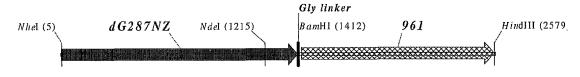


FIGURE 14G — ΔG983-ORF46.1



10/13

FIGURE 14H — ΔG983-741



FIGURE 14I — ΔG983-961



FIGURE 14J — ΔG983-961c



FIGURE 14K — ΔG741-961



FIGURE 14L — ΔG741-961c



11/13

FIGURE 14M — ΔG741-983

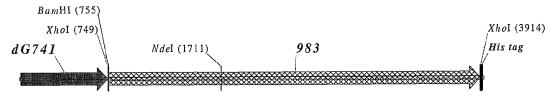


FIGURE 14N — ΔG741-ORF46.1



FIGURE 140 — ORF46.1-741

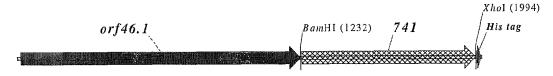


FIGURE 14P — ORF46.1-961

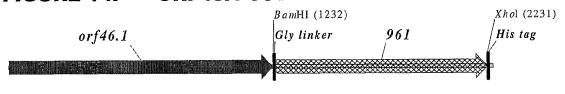
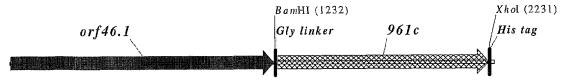


FIGURE 14Q — ORF46.1—961c



12/13

FIGURE 14R — 961-ORF46.1

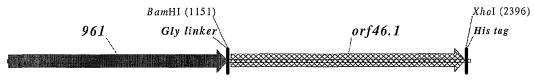


FIGURE 14S -- 961-741

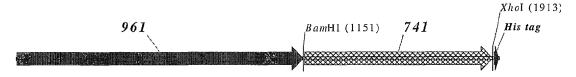


FIGURE 14T — 961-983

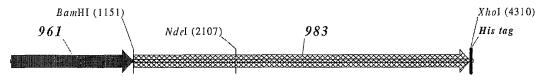
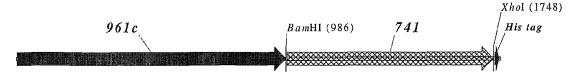


FIGURE 14U — 961c-ORF46.1



FIGURE 14V — 961c-741



13/13

FIGURE 14W — 961c-983

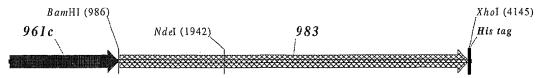


FIGURE 14X — 961cL-ORF46.1

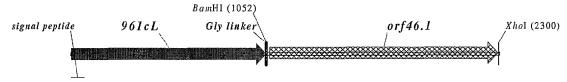


FIGURE 14Y — 961cL-741

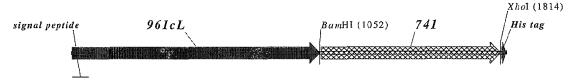


FIGURE 14Z — 961cL-983

